# Glauber dynamics on trees: Boundary conditions and mixing time[†]

Fabio Martinelli[‡] Alistair Sinclair[§] Dror Weitz[¶]

February 1, 2008



**Abstract**

We give the first comprehensive analysis of the effect of boundary conditions on the mixing time of the Glauber dynamics in the so-called *Bethe approximation*. Specifically, we show that spectral gap and the log-Sobolev constant of the Glauber dynamics for the Ising model on an $n$-vertex regular tree with $(+)$-boundary are bounded below by a constant independent of $n$ at all temperatures and all external fields. This implies that the mixing time is $O(\log n)$ (in contrast to the free boundary case, where it is not bounded by any fixed polynomial at low temperatures). In addition, our methods yield simpler proofs and stronger results for the spectral gap and log-Sobolev constant in the regime where there are multiple phases but the mixing time is insensitive to the boundary condition. Our techniques also apply to a much wider class of models, including those with hard-core constraints like the antiferromagnetic Potts model at zero temperature (proper colorings) and the hard–core lattice gas (independent sets).

[†]An extended abstract of this paper appeared under the title "The Ising model on trees: Boundary conditions and mixing time" in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, October 2003.

[‡]Department of Mathematics, University of Roma Tre, Largo San Murialdo 1, 00146 Roma, Italy. Email: `martin@mat.uniroma3.it`. This work was done while this author was visiting the Departments of EECS and Statistics, University of California, Berkeley, supported in part by a Miller Visiting Professorship.

[§]Computer Science Division, University of California, Berkeley, CA 94720-1776, U.S.A. Email: `sinclair@cs.berkeley.edu`. Supported in part by NSF Grant CCR-0121555 and DARPA cooperative agreement F30602-00-2-0601.

[¶]Computer Science Division, University of California, Berkeley, CA 94720-1776, U.S.A. Email: `dror@cs.berkeley.edu`. Supported in part by NSF Grant CCR-0121555.

# 1 Introduction

In this paper we will analyze the influence of boundary conditions on the Glauber dynamics for discrete spin models on a regular rooted tree. Although in what follows we will focus for simplicity on the well known Ising model, our techniques also apply to other models, not necessarily ferromagnetic and with hard-core constraints.

In the Ising model on a finite graph $G = (V, E)$, a configuration $\sigma = (\sigma_x)$ consists of an assignment of $\pm 1$-values, or "spins", to each vertex (or "site") of $V$. The probability of finding the system in configuration $\sigma \in \{\pm 1\}^V \equiv \Omega_G$ is given by the *Gibbs distribution*

$$\mu_G(\sigma) \propto \exp\Big(\beta \sum\nolimits_{xy \in E} \sigma_x \sigma_y + \beta h \sum\nolimits_{x \in V} \sigma_x\Big), \tag{1}$$

where $\beta \geq 0$ is the inverse temperature and $h$ the external field. Boundary conditions can also be taken into account by fixing the spin values at some specified "boundary" vertices of $G$; the term *free boundary* is used to indicate that no boundary condition is specified.

In the classical Ising model, $G = G_n$ is a cube of side $n^{1/d}$ in the $d$-dimensional Cartesian lattice $\mathbb{Z}^d$, and in this case the phase diagram in the thermodynamic limit $G_n \uparrow \mathbb{Z}^d$ is quite well understood (see, e.g., [15, 38] for more background).

While the classical theory focused on static properties of the Gibbs measure, in the last decade the emphasis has shifted towards dynamical questions with a computational flavor. The key object here is the *Glauber dynamics*, a (discrete– or continuous–time) Markov chain on the set of spin configurations $\Omega_G$ in which each spin $\sigma_x$ flips its value with a rate that depends on the current configuration of the neighboring spins of $x$, and which satisfy the detailed balance condition w.r.t to the Gibbs measure $\mu_G$ (see Section 2 for more details).

The Glauber dynamics is much studied for two reasons: firstly, it is the basis of Markov chain Monte Carlo algorithms, widely used in computational physics for sampling from the Gibbs distribution; and secondly, it is a plausible model for the actual evolution of the underlying physical system towards equilibrium. In both contexts, one of the central questions is to determine the *mixing time*, i.e., the time until the dynamics is close to its stationary distribution.

As is well known (see e.g. [36]), the approach to stationarity of a reversible Markov chain with Markov generator $\mathcal{L}$ and reversible measure $\pi$ can be succesfully studied by analyzing two key quantities: the *spectral gap* and the *logarithmic Sobolev constant* of the pair $(\mathcal{L}, \pi)^\dagger$. The first of these measures the rate of the exponential decay as $t \to \infty$ of the variance $\mathrm{Var}_\pi(e^{t\mathcal{L}} f)$ computed with respect to the invariant measure $\pi$, while the second measures instead the rate of decay of the relative entropy of $e^{t\mathcal{L}} f$ w.r.t $\pi$ (see, e.g., [1]). Advances in statistical physics over the past decade have led to remarkable connections between these two quantities and the occurence of a phase transition (see, e.g., [40, 30, 29, 9, 28, 26]). As an example, on finite $n$-vertex squares with free boundary in the 2-dimensional lattice $\mathbb{Z}^2$, when $h = 0$ and $\beta$ is smaller than the critical value $\beta_c$, the spectral gap and the logarithmic Sobolev constant are $\Omega(1)$ (*i.e.* bounded away from zero uniformly in $n$), while for $\beta > \beta_c$ they are both exponentially small in $\sqrt{n}$.

One of the most interesting and difficult questions left open by the above and related results is the *influence of boundary conditions* on the spectral gap and the log-Sobolev constant when $h = 0$ and $\beta > \beta_c$. It has been conjectured that, in the presence of an all-$(+)$ boundary, the relaxation process is driven by the mean–curvature motion of interfaces separating droplets of the $(-)$-phase inside the $(+)$-phase, and therefore the mixing time should be polynomial in $n$ (most likely $n^{2/d} \log n$) [7, 14]. In particular it has been argued that the spectral gap for the pure phases in high enough dimension should be $\Omega(1)$. Proving results of this kind has proved very elusive,

$^\dagger$Unfortunately the definition of the logarithmic Sobolev constant is not constant in the literature. The ambiguity arises because there are two definitions, one the inverse of the other. The definition used in this paper is the one that puts the logarithmic Sobolev constant and the spectral gap on the same footing.

and the only (presumably sharp) available bounds are *upper bounds* on the spectral gap and the logarithmic Sobolev constant [7].

In this paper we prove a strong version of the above conjecture in what is known in statistical physics as the *Bethe approximation*, namely when the lattice $\mathbb{Z}^d$ is replaced by a regular tree. Among other results, we show that the spectral gap of the Glauber dynamics for the Ising model on a tree with a $(+)$-boundary condition on its leaves is $\Omega(1)$ at all temperatures and all values of the external field, and further that the same holds for the logarithmic Sobolev constant. Notice that, with a free boundary, $\beta$ large and $h = 0$, both quantities tend to zero as $1/n^a$ and the exponent $a$ grows arbitrarily large as $\beta \to \infty$ [3].

Ours is apparently the first result that quantifies the effect of boundary conditions on Glauber dynamics in an interesting scenario. We stress that, while the tree is simpler in many respects than $\mathbb{Z}^d$ due to the lack of cycles, in other respects it is more complex due to the large boundary: e.g., it exhibits a "double phase transition," and the critical field at low temperature is non–zero (see below). In the next subsection, we briefly describe the Ising model on trees before stating our results in more detail.

## 1.1 The Ising model on trees

Fix $b \geq 2$ and let $\mathbb{T}^b$ denote the infinite $b$-ary tree. The Ising model on $\mathbb{T}^b$ is known [15, 24] to have a phase diagram in the $(h, \beta)$ plane quite different from that on the cubic lattice $\mathbb{Z}^d$ (see Fig. 1), and has recently received a lot of attention as the canonical example of a statistical physics model on a "non-amenable" graph (i.e., one whose boundary is of comparable size to its volume) — see, e.g., [6, 19, 13, 37, 22, 3, 5].

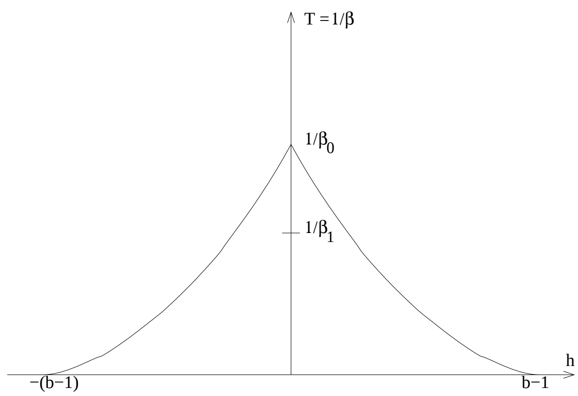


Figure 1: The critical field $h_c(\beta)$. The Gibbs measure is unique above the curve.

Let us first discuss the behavior on the line $h = 0$. There is a first critical value $\beta_0 = \frac{1}{2}\log\left(\frac{b+1}{b-1}\right)$, marking the dividing line between uniqueness and non-uniqueness of the Gibbs measure. Then, in sharp contrast to the model on $\mathbb{Z}^d$, there is a second critical point $\beta_1 = \frac{1}{2}\log\left(\frac{\sqrt{b}+1}{\sqrt{b}-1}\right)$ which is often referred to as the "spin-glass critical point" [10]. This second critical point is such that, in the "intermediate temperature" region $\beta_0 < \beta \leq \beta_1$, the $(+)$- and $(-)$-boundary conditions exert arbitrarily long-range influence on the spin at root of the tree and hence give rise to different Gibbs measures, but "typical" boundary conditions (i.e., chosen from the infinite volume Gibbs measure with free boundary) do not. Another way to phrase this peculiar behavior is that the Gibbs measure constructed via a free boundary is *extremal* for all $\beta \leq \beta_1$ (see [6, 19, 20, 3] and also [13, 33, 34] for an analysis in the context of "bit reconstruction problems" for noisy data transmission).

Let us now examine what happens when an external field $h$ is added to the system. It turns out that for all $\beta > \beta_0$, there is a critical value $h = h_c(\beta) > 0$ of the field such that the Gibbs measure is not unique when $|h| \leq h_c$, and is unique when $|h| > h_c$. (When $\beta \leq \beta_0$ the Gibbs measure is

unique for all $h$, and $h_c$ is defined to be zero.) In the presence of a $(+)$-boundary, the Ising model on the tree with external field $h = -h_c$ is rather analogous to the classical case of $\mathbb{Z}^d$ with zero field. Both models share the following two properties: firstly, the Gibbs measure is sensitive to the choice of boundary condition, and secondly, adding an arbitrarily small negative field causes the Gibbs measure to become insensitive to the boundary condition (i.e., unique in the thermodynamic limit).

Finally we remark that the concentration properties of the Gibbs measure for $\beta > \beta_0$, $h \geq -h_c$ and $(+)$-boundary are very different from those on $\mathbb{Z}^d$. In the latter case, along the line of first order phase transition, the (negative) large deviations for the bulk magnetization are related to the appearance of a Wulff droplet of the opposite phase and are depressed by a negative exponential in the *surface* of the droplet (see, e.g., [11]). Here instead, for any value of $(\beta, h)$ they are always depressed by a negative exponential in the *volume* of the excess negative spins (the phenomenon of "rigidity of the critical phases" [5]).

The Glauber dynamics for the Ising model on trees has also been studied. In a recent paper [3], it is shown that the associated spectral gap (see (7) for a precise definition) with zero external field and free boundary on a complete $b$-ary tree $T$ with $n$ vertices is $\Omega(1)$ at high and intermediate temperatures (i.e., when $\beta < \beta_1$)[‡]. Moreover, at the critical point $\beta = \beta_1$ the same spectral gap is bounded above by $c/\log n$, and as soon as $\beta > \beta_1$ it becomes smaller than $c/n^{a(\beta)}$, with $a(\beta) \uparrow \infty$ as $\beta \to \infty$. Thus the critical point $\beta = \beta_1$ is reflected in the dynamics by an abrupt jump in the behavior of the spectral gap as a function of the size of the tree $T$. Finally, also in [3], it is proved that the spectral gap for arbitrary fixed $\beta, h$ and boundary condition can never shrink to zero faster than an inverse polynomial in $n$. Again such a result should be compared to the lattice case where it is known that the spectral gap for a cube with $n$ sites can be exponentially small in the surface $n^{(d-1)/d}$.

## 1.2 Main results and techniques

Our first main result is a detailed analysis of the spectral gap of the Glauber dynamics in different regions of the phase diagram. The main novelty here is that we are able for the first time to prove a sharp result in the region where the spectral gap is highly sensitive to the boundary condition.

**Theorem 1.1** *In both of the following situations, the spectral gap of the Glauber dynamics on a complete $b$-ary tree $T$ with $n$ vertices is $\Omega(1)$:*

(i) *the boundary condition is arbitrary, and either $\beta < \beta_1$ (with $h$ arbitrary), or $|h| > h_c(\beta)$ (with $\beta$ arbitrary);*

(ii) *the boundary condition is $(+)$ and $\beta, h$ are arbitrary.*

**Remark:** On $\mathbb{Z}^d$ not much is known about the spectral gap when $\beta > \beta_c$, $h = 0$ and the boundary condition is $(+)$, the notable exception being that of $\mathbb{Z}^2$ where it has been recently proved [7] that the spectral gap in a square with $n$ sites shrinks to zero at least as $1/\sqrt{n}$ (neglecting logarithmic corrections). The best known lower bounds are significantly weaker [28]. In high enough dimensions ($d \geq 3$) it has been conjectured (see [14] and [7]) that the spectral gap should stay bounded away from zero uniformly in $n$. The above theorem can be looked upon as evidence in favor of this conjecture.

In our second main result we extend our analysis to the more delicate and difficult logarithmic Sobolev constant (see (7) for a precise definition).

[‡]Actually the arguments in [3] prove that the gap is $\Omega(1)$ for any $\beta < \beta_1$, arbitrary boundary condition and any external field. Their argument, together with some monotonicity properties specific to the Ising model [35], implies a mixing time of $O(\log n)$. Thus, although for $\beta_0 < \beta < \beta_1$ there exist several Gibbs measures, the mixing time of the Glauber dynamics is *insensitive* to the boundary condition.

**Theorem 1.2** *In the same situations as in Theorem 1.1, the logarithmic Sobolev constant of the Glauber dynamics on a complete $b$-ary tree $T$ with $n$ vertices is $\Omega(1)$.*

As a corollary we obtain that, in the situations of Theorems 1.1 and 1.2, the Glauber dynamics mixes (in a very strong sense) in time $O(\log n)$.

**Remarks:**

(i) In $\mathbb{Z}^d$ with $(+)$-boundary condition, $\beta$ large and zero external field the logarithmic Sobolev constant in a cube with $n$ sites is always smaller than $n^{-2/d}$, neglecting logarithmic corrections [7], in agreement with heuristic predictions based on mean–curvature motion of phases interfaces.

(ii) We also prove (see Theorem 5.7) an additional result which shows that, for an arbitrary nearest-neighbor spin system on a tree, as soon as the spectral gap is $\Omega(1)$ then the logarithmic Sobolev constant cannot shrink faster than $(c \log n)^{-1}$. This means that, even when a constant lower bound is known for the gap but not for log-Sobolev, one can deduce a mixing time of $O((\log n)^2)$. While we do not require this fact to derive the results of this paper, we believe it may be of interest for other models on trees.

In order to better appreciate Theorem 1.2, one should keep in mind that for general finite range, translation invariant, compact spin models on $\mathbb{Z}^d$, if there exists an infinite volume Gibbs measure $\mu$ with a positive logarithmic Sobolev constant, then the system is necessarily in the uniqueness region and $\mu$ has exponentially decaying correlations [41][§]. We also recall (see, e.g., [25]) that when the log-Sobolev constant is bounded away from zero one can derive very strong (Gaussian–like) concentration properties of the corresponding Gibbs measure, such as those proved in [5].

We now proceed to sketch some of our techniques and point out the main technical innovations.

Our analysis of both the log-Sobolev constant and the spectral gap rests on certain *spatial mixing conditions* that can be stated as follows. Let $f$ be a function of the spin configuration that *does not depend* on the spins in the first $\ell$ levels of the tree starting from the root $r$, and let $\mu(f \mid \sigma_r)$ be the projection of $f$ onto the spin $\sigma_r$ at the root. If the variance (respectively, the entropy) under the Gibbs measure $\mu$ of $\mu(f \mid \sigma_r)$ decays fast enough with the depth $\ell$, then we show by a unified argument how to deduce a bound of $\Omega(1)$ on the spectral gap (respectively, the log-Sobolev constant). Crucially, in contrast to previous approaches we do not require the above decay to hold in arbitrary environments, but only for the Gibbs measure $\mu$ under consideration. This opens up the possibility that the condition holds for some boundary conditions and not for others (with the same values of temperature and external field). We also prove the converse, thus showing the that our mixing conditions are in fact equivalent to the required bounds on the spectral gap and log–Sobolev constants.

This analysis has several advantages over previous ones [3, 35]: it is more direct, applies also when there is an external field, and applies to general nearest-neighbor spin systems on trees.

The second main ingredient of the paper is establishing the above spatial mixing conditions in the scenarios of interest described in the above two theorems. This is done via a rather simple and novel coupling technique for the case of the variance. Such a technique provides, along the way, a new and really elementary proof of the extremality of the Gibbs measure with free boundary below $\beta_1$.

Surprisingly, we are also able to exploit the same coupling technique (via strong concentration properties of the Gibbs measure) to establish the entropy mixing condition. Thus in terms of the coupling analysis our conditions for variance and entropy mixing are essentially the same.

[§]A close look at the proof in [41] reveals that the same is true for any infinite, locally finite, bounded degree graph such that the volume of any ball of radius $\ell$ grows sub–exponentially in $\ell$.

Finally, we mention that our results actually hold (with suitable modifications) for a much wider class of spin systems on trees than just the Ising model, including the Potts model and models with hard constraints such as the zero-temperature antiferromagnetic Potts model (proper colorings) and the hard-core lattice gas model (independent sets). We briefly outline some of these extensions at the end of the paper; full details can be found in a companion paper [31].

The remainder of the paper is organized as follows. In Section 2 we give some basic definitions and notation. Then in Section 3 we define the spatial mixing conditions and relate them to the spectral gap and log-Sobolev constant. The mixing conditions in the scenarios of interest for the spectral gap and the log-Sobolev constant are verified in Sections 4 and 5 respectively. Finally, in Section 6 we mention some extensions of our results to other models of interest. The proofs of some technical lemmas omitted from the main text are collected in a supplement, Section 7.

## Acknowledgments

F. Martinelli would like to thank the Miller Institute, the Dept. of Statistics and the Dept. of EECS of the University of California at Berkeley for financial support and warm hospitality. We also wish to thank E. Mossel and Y. Peres for very interesting discussions about reconstruction on trees and related topics.

## 2 Preliminaries

### 2.1 Gibbs distributions on trees

For $b \geq 2$, let $\mathbb{T}^b$ denote the infinite, rooted $b$-ary tree (in which every vertex has $b$ children). We will be concerned with (complete) finite subtrees $T$ of $\mathbb{T}^b$; if $T$ has depth $m$ then it has $n = (b^{m+1}-1)/(b-1)$ vertices, and its *boundary* $\partial T$ consists of the children (in $\mathbb{T}^b$) of its leaves, i.e., $|\partial T| = b^{m+1}$. We identify subgraphs of $T$ with their vertex sets, and write $E(A)$ for the edges within a subset $A$, and $\partial A$ for the boundary of $A$ (i.e., the neighbors of $A$ in $(T \cup \partial T) \setminus A$).

Fix an Ising spin configuration $\tau$ on the infinite tree $\mathbb{T}^b$. We denote by $\Omega_T^\tau$ the set of (finite) spin configurations $\sigma \in \{\pm 1\}^{T \cup \partial T}$ that agree with $\tau$ on $\partial T$; thus $\tau$ specifies a *boundary condition* on $T$. Usually we abbreviate $\Omega_T^\tau$ to $\Omega$. For any $\eta \in \Omega$ and any subset $A \subseteq T$, we denote by $\mu_A^\eta$ the Gibbs distribution over $\Omega$ conditioned on the configuration outside $A$ being $\eta$: i.e., if $\sigma \in \Omega$ agrees with $\eta$ outside $A$ then

$$\mu_A^\eta(\sigma) \propto \exp\Big[\beta\big(\sum\nolimits_{xy \in E(A \cup \partial A)} \sigma_x \sigma_y + h \sum\nolimits_{x \in A} \sigma_x\big)\Big],$$

where $\beta$ is the inverse temperature and $h$ the external field. We define $\mu_A^\eta(\sigma) = 0$ otherwise. In particular, when $A = T$, $\mu_T^\tau$ is simply the Gibbs distribution on the whole of $T$ with boundary condition $\tau$; we abbreviate $\mu_T^\tau$ to $\mu$.

For a function $f : \Omega \to \mathbb{R}$ we denote by $\mu_A^\eta(f) = \sum_{\sigma \in \Omega} \mu_A^\eta(\sigma) f(\sigma)$ the *expectation* of $f$ w.r.t. the distribution $\mu_A^\eta$. It will be convenient to view $\mu_A^\eta(f)$ as a function of $\eta$, defined by $\mu_A(f)(\eta) = \mu_A^\eta(f)$, the *conditional expectation* of $f$. Note that $\mu_A(f)$ is a function from $\Omega$ to $\mathbb{R}$ but depends only on the configuration outside $A$. We write $\mathrm{Var}_A^\eta(f) = \mu_A^\eta(f^2) - \mu_A^\eta(f)^2$ and (for $f \geq 0$) $\mathrm{Ent}_A^\eta(f) = \mu_A^\eta(f \log f) - \mu_A^\eta(f) \log \mu_A^\eta(f)$ for the *variance* and *entropy* of $f$ respectively w.r.t. $\mu_A^\eta$. Note that $\mathrm{Var}_A^\eta(f) = 0$ iff, conditioned on the configuration outside $A$ being $\eta$, $f$ does not depend on the configuration inside $A$. The same holds for $\mathrm{Ent}_A^\eta(f)$. In case $A = T$ we use the abbreviations $\mu(f), \mathrm{Var}(f)$ and $\mathrm{Ent}(f)$.

We record here some basic properties of variance and entropy that we use throughout the paper: **(i)** For $B \subseteq A \subseteq T$,

$$\mathrm{Var}_A^\eta(f) = \mu_A^\eta[\mathrm{Var}_B(f)] + \mathrm{Var}_A^\eta[\mu_B(f)]. \tag{2}$$

This equation expresses a decomposition of the variance into the local conditional variance in $B$ and the variance of the projection outside $B$.
**(ii)** If $A = \bigcup_i A_i$ for disjoint $A_i$, and the Gibbs distribution $\mu_A^\eta$ is the product of its marginals over the $A_i$, then for any function $f$,

$$\mathrm{Var}_A^\eta(f) \le \sum_i \mu_A^\eta[\mathrm{Var}_{A_i}(f)]. \tag{3}$$

**(iii)** For any two subsets $A, B \subseteq T$ such that $(\partial A) \cap B = \emptyset$, and for any function $f$,

$$\mu[\mathrm{Var}_A(\mu_B(f))] \le \mu[\mathrm{Var}_A(\mu_{A\cap B}(f))]. \tag{4}$$

Properties (ii) and (iii) are consequences of the fact that variance w.r.t. a fixed measure is a convex functional.
All three properties (i), (ii) and (iii) also hold with $\mathrm{Var}$ replaced by $\mathrm{Ent}$.

## 2.2 The Glauber dynamics

The *Glauber dynamics* on $T$ with boundary conditions $\tau$ is the continuous time Markov chain on $\Omega = \Omega_T^\tau$ with Markov generator $\mathcal{L} \equiv \mathcal{L}_T^\tau$ given by

$$(\mathcal{L}f)(\sigma) = \sum_{x\in T} c_x(\sigma)[f(\sigma^x) - f(\sigma)], \tag{5}$$

where $\sigma^x$ denotes the configuration obtained from $\sigma$ by flipping the spin at the site $x$, and $c_x(\sigma)$ denotes the flip rate at $x$. Although all our results apply to any choice of finite–range, uniformly positive and bounded flip rates satisfying the detailed balance condition w.r.t. the Gibbs measure, for simplicity in the sequel we will work with a specific choice known as the *heat-bath* dynamics:

$$c_x(\sigma) = \mu_{\{x\}}^\sigma(\sigma^x) = \frac{1}{1+w_x(\sigma)}, \quad \text{where} \quad w_x(\sigma) = \exp\bigl[2\beta\sigma_x(\sum_{xy\in E}\sigma_y + h)\bigr].$$

It is a well-known fact (and easily checked) that the Glauber dynamics is ergodic and reversible w.r.t. the Gibbs distribution $\mu = \mu_T^\tau$, and so converges to the stationary distribution $\mu$. The rate of convergence is often measured using two concepts from functional analysis: the *spectral gap* and the *logarithmic Sobolev constant*. For a function $f : \Omega \to \mathbb{R}$, define the *Dirichlet form* of $f$ associated with the generator $\mathcal{L}$ by

$$\mathcal{D}(f) := \tfrac{1}{2}\sum_x \mu\bigl(c_x\bigl[f(\sigma^x) - f(\sigma)\bigr]^2\bigr) = \sum_x \mu(\mathrm{Var}_{\{x\}}(f)). \tag{6}$$

(The l.h.s. here is the general definition for any choice of the flip rates $c_x$; the last equality holds when specializing to the case of the heat-bath dynamics.) The *spectral gap* $c_{\mathrm{gap}}(\mu)$ and the *logarithmic Sobolev constant* $c_{\mathrm{sob}}(\mu)$ of the chain are then defined by

$$c_{\mathrm{gap}}(\mu) = \inf_f \frac{\mathcal{D}(f)}{\mathrm{Var}(f)}; \qquad c_{\mathrm{sob}}(\mu) = \inf_{f\ge 0} \frac{\mathcal{D}(\sqrt{f}\,)}{\mathrm{Ent}(f)}, \tag{7}$$

where the infimum in each case is over non-constant functions $f$.

As is well known, these two quantities measure the rate of exponential decay as $t \to \infty$ of the variance and relative entropy respectively (see, e.g., [36]). The quantity $c_{\mathrm{gap}}$ also has a natural interpretation as the smallest positive eigenvalue of $-\mathcal{L}$.

We make the following important note. When discussing the asymptotics of $c_{\mathrm{sob}}$ (or $c_{\mathrm{gap}}$) for a fixed boundary condition $\tau$, we think of the infinite sequence of Gibbs distributions $\{\mu_T^\tau\}$, where $T$

ranges over all finite complete subtrees of $\mathbb{T}^b$. In particular, when we say that $c_{\rm sob}(\mu) = c_{\rm sob}(\mu_T^\tau) = \Omega(1)$ we mean that there exists a finite constant $C > 0$ such that for every $T$ (or equivalently, for every $\mu \in \{\mu_T^\tau\}$), $c_{\rm sob}(\mu) \geq 1/C$.

We close this section by recalling some well-known relationships between the above constants and certain notions of mixing time of the Glauber dynamics. Define $h_t^\sigma(\eta) = \frac{P_t(\sigma,\eta)}{\mu(\eta)}$, where $P_t(\sigma,\eta) := e^{t\mathcal{L}}(\sigma,\eta)$ is the transition kernel at time $t$. Then, for $1 \leq p \leq \infty$, define

$$T_p := \min\Big\{t > 0 : \sup_\sigma \|h_t^\sigma - 1\|_p \leq \frac{1}{e}\Big\} \tag{8}$$

where $\|f\|_p$ denotes the $L^p(\Omega,\mu)$ norm of $f$. The time $T_1$ is usually called simply the *mixing time* of the chain. Standard results relating $T_p$ to the spectral gap and log-Sobolev constant (see, e.g., [36]), when specialized to the Glauber dynamics, yield the following:

**Theorem 2.1** *On an $n$-vertex $b$-ary tree $T$ with boundary condition $\tau$,*

(i) $c_{\rm gap}(\mu)^{-1} \leq T_1 \leq c_{\rm gap}(\mu)^{-1} \times C_1 n$;

(ii) $c_{\rm gap}(\mu)^{-1} \leq T_2 \leq c_{\rm sob}(\mu)^{-1} \times C_2 \log n$,

*where $\mu = \mu_T^\tau$ and $C_1, C_2$ are constants depending only on $b, \beta$ and $h$.* □

Finally, we note that our choice of the heat-bath dynamics is not essential. Since changing to any other reversible local update rule (e.g., the Metropolis rule) affects $c_{\rm sob}$ and $c_{\rm gap}$ by at most a constant factor, our analysis applies to any choice of Glauber dynamics.

# 3 Spatial mixing conditions for spectral gap and log-Sobolev

In this section we define a certain *spatial mixing condition* (i.e., a form of weak dependence between the spin at a site and the configuration far from that site) for a Gibbs distribution $\mu$, and prove that this condition implies that $c_{\rm gap}(\mu) = \Omega(1)$. An analogous condition implies that $c_{\rm sob}(\mu) = \Omega(1)$. Our spatial mixing conditions have two main advantages over those used previously: first, the conditions for the spectral gap and the log-Sobolev constant are identical in form, allowing a uniform treatment; second, and more importantly, they are *measure-specific*, i.e., they may hold for the Gibbs distribution induced by some specific boundary configuration while not holding for other boundary configurations. Hence, the conditions are sensitive enough to show rapid mixing for specific boundaries even though the mixing time with other boundaries is slow for the same choice of temperature and external field. We also note that the results of this section hold not just for the Ising model but for any nearest-neighbor interaction model on a tree.

## 3.1 Reduction to block analysis

Before presenting the main result of this section, we need some more definitions and background. For each site $x \in T$, let $B_{x,\ell} \subseteq T$ denote the subtree (or "block") of height $\ell - 1$ rooted at $x$, i.e., $B_{x,\ell}$ consists of $\ell$ levels. (If $x$ is $k < \ell$ levels from the bottom of $T$ then $B_{x,\ell}$ has only $k$ levels.) In what follows we will think of $\ell$ as a suitably large constant. By analogy with expression (6) for the Dirichlet form, let $\mathcal{D}_\ell(f) \equiv \sum_{x\in T} \mu[{\rm Var}_{B_{x,\ell}}(f)]$ denote the local variation of $f$ w.r.t. the blocks $\{B_{x,\ell}\}$. A straightforward manipulation (see, e.g., [28], keeping in mind that each site belongs to at most $\ell$ blocks) shows that $c_{\rm gap}$ can be bounded as follows:

$$c_{\rm gap}(\mu) \geq \frac{1}{\ell} \cdot \inf_f \frac{\mathcal{D}_\ell(f)}{{\rm Var}(f)} \cdot \min_{\eta,x} c_{\rm gap}(\mu^\eta_{B_{x,\ell}}). \tag{9}$$

As before, the infimum is taken over non-constant functions (and henceforth we omit explicit mention of this). The importance of (9) is that $\min_{\eta,x} c_{\rm gap}(\mu^{\eta}_{B_{x,\ell}})$ depends only on the size of $B_{x,\ell}$ and $\beta$, but not on the size of $T$; in fact, it is at least $\Omega(e^{-c(b,\beta)\cdot\ell})$ [3]. Therefore, in order to show that $c_{\rm gap}$ is bounded by a constant independent of the size of $T$, it is enough to show that, for some finite $\ell$, $\mathrm{Var}(f) \le \mathrm{const} \times \mathcal{D}_\ell(f)$ for all functions $f$. This is what we will show below, under the relevant spatial mixing condition. As a side remark, notice that $\inf_f \frac{\mathcal{D}_\ell(f)}{\mathrm{Var}(f)}$ is exactly the spectral gap of the Glauber dynamics based on flipping *blocks* $B_{x,\ell}$, rather than single sites $x$.

An identical manipulation yields an analogous bound for the log-Sobolev constant. For a non-negative function $f$, let $\mathcal{E}_\ell(f) \equiv \sum_{x\in T}\mu[\mathrm{Ent}_{B_{x,\ell}}(f)]$. Then

$$c_{\rm sob}(\mu) \ge \frac{1}{\ell}\cdot \inf_{f\ge 0}\frac{\mathcal{E}_\ell(f)}{\mathrm{Ent}(f)}\cdot \min_{\eta,x} c_{\rm sob}(\mu^{\eta}_{B_{x,\ell}}). \tag{10}$$

Hence to bound $c_{\rm sob}(\mu)$ it suffices to show that, for some constant $\ell$, $\mathrm{Ent}(f) \le \mathrm{const}\times\mathcal{E}_\ell(f)$ for all $f \ge 0$.

### 3.2 Spatial mixing

We are now ready to state our spatial mixing conditions, first for the variance and then for the entropy. For $x \in T$, write $T_x$ for the subtree rooted at $x$, and $\widetilde{T_x}$ for $T_x \setminus \{x\}$, the subtree $T_x$ excluding its root.

**Definition 3.1 [Variance Mixing]** *We say that $\mu = \mu^\tau_T$ satisfies $\mathrm{VM}(\ell,\varepsilon)$ if for every $x \in T$, any $\eta \in \Omega^\tau_T$ and any function $f$ that does not depend on $B_{x,\ell}$, the following holds:*

$$\mathrm{Var}^{\eta}_{T_x}[\mu_{\widetilde{T_x}}(f)] \;\le\; \varepsilon\cdot \mathrm{Var}^{\eta}_{T_x}(f).$$

Let us briefly discuss the above condition. Essentially, $\varepsilon = \varepsilon(\ell)$ gives the rate of decay with distance $\ell$ of point-to-set correlations. To see this, note that the l.h.s. $\mathrm{Var}^{\eta}_{T_x}[\mu_{\widetilde{T_x}}(f)]$ is the variance of the *projection* of $f$ onto the root $x$ of $T_x$, which is at distance $\ell$ from the sites on which $f$ depends. It is also worth noting that the required uniformity in $\eta$ in VM is not very restrictive: since the distribution $\mu^{\eta}_{T_x}$ depends only on the restriction of $\eta$ to the boundary of $T_x$, and since $\eta \in \Omega^\tau_T$ (i.e., $\eta$ agrees with $\tau$ on $\partial T$ and therefore on the bottom boundary of $T_x$), the only freedom left in choosing $\eta$ is in choosing the spin of the parent of $x$. Thus, VM is essentially a property of the distribution induced by the boundary condition $\tau$. It is this lack of uniformity (i.e., the fact that we need not verify VM for other boundary conditions) that makes it flexible enough for our applications.

As the following theorem states, if $\mathrm{VM}(\ell,\varepsilon)$ holds with $\varepsilon \approx \frac{1}{2\ell}$, then we get a lower bound on $c_{\rm gap}$:

**Theorem 3.2** *For any $\ell$ and $\delta > 0$, if $\mu$ satisfies $\mathrm{VM}(\ell,(1-\delta)/2(\ell+1-\delta))$ then $\mathrm{Var}(f) \le \frac{3}{\delta}\cdot\mathcal{D}_\ell(f)$ for all $f$. In particular, if VM with the above parameters holds for some fixed $\ell$ and $\delta > 0$, for all $\mu = \mu^\tau_T$ with $T$ a full subtree, then $c_{\rm gap}(\mu) = \Omega(1)$. Conversely, if $c_{\rm gap}(\mu) = \Omega(1)$ then for all $T$, $\mu^\tau_T$ satisfies $\mathrm{VM}(\ell, ce^{-\vartheta\ell})$ for some constants $c,\vartheta > 0$ and all $\ell$.*

**Remark:** The second part of the theorem was already proved in [3], where it was shown that for general nearest-neighbor spin systems on any bounded degree graph, if $c_{\rm gap}(\mu)$ is bounded independently of $n$ then $\mu$ exhibits an exponential decay of point-to-set correlations (i.e., $\mathrm{VM}(\ell, c\exp(-\vartheta\ell))$ holds for all $\ell$). The authors of [3] posed the question of whether the converse is also true. Theorem 3.2 (which holds for general nearest-neighbor spin systems on a tree) answers this question affirmatively when the graph is a tree. In fact, as is apparent from the above theorem, the decay of point-to-set correlations on a tree is either slower than linear or exponentially fast.

The analogous mixing condition for entropy and the log-Sobolev constant is the following:

**Definition 3.3 [Entropy Mixing]** *We say that $\mu = \mu_T^\tau$ satisfies $\mathrm{EM}(\ell, \varepsilon)$ if for every $x \in T$, any $\eta \in \Omega_T^\tau$ and any non-negative function $f$ that does not depend on $B_{x,\ell}$, the following holds:*

$$\mathrm{Ent}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \;\le\; \varepsilon \cdot \mathrm{Ent}_{T_x}^\eta(f).$$

Before stating the analog of Theorem 3.2 relating $c_{\mathrm{sob}}$ to EM, we need to define one more constant. Let $p_{\min} = \min_{x,s,\eta\in\Omega_T^\tau} \mu_{T_x}^\eta(\sigma_x = s)$, where $s$ ranges over $\{+,-\}$; i.e., $p_{\min}$ is the minimum probability of any spin value at any site with any boundary condition. It is easy to see that $p_{\min} \ge \frac{1}{2}e^{-2\beta(b+|h|)}$, a constant depending only on $b, \beta, h$.

**Theorem 3.4** *For any $\ell$ and $\delta > 0$, if $\mu$ satisfies $\mathrm{EM}(\ell, [(1-\delta)p_{\min}/(\ell+1-\delta)]^2)$ then $\mathrm{Ent}(f) \le \frac{2}{\delta}\cdot\mathcal{E}_\ell(f)$ for all $f \ge 0$. In particular, if EM with the above parameters holds for some fixed $\ell$ and $\delta > 0$, for all $\mu = \mu_T^\tau$ with $\tau$ fixed and $T$ an arbitrary full subtree, then $c_{\mathrm{sob}}(\mu) = \Omega(1)$. Conversely, if $c_{\mathrm{sob}}(\mu) = \Omega(1)$ then for all $T$, $\mu_T^\tau$ satisfies $\mathrm{EM}(\ell, ce^{-\vartheta\ell})$ for some constants $c, \vartheta > 0$ and all $\ell$.*

In order to prove Theorems 3.2 and 3.4 it is convenient to work with spatial mixing conditions that are somewhat more involved than VM and EM. The main difference is that we want to allow for functions that may depend on $B_{x,\ell}$ (the first $\ell$ levels of $T_x$) and thus need to introduce a term for this dependency. The modified conditions express the property that the variance (entropy) of the projection of any function $f$ onto the root $x$ of $T_x$ can be bounded up to a constant factor by the local variance (entropy) of $f$ in $B_{x,\ell}$, plus a negligible factor times the local variance (entropy) of $f$ in $\widetilde{T_x}$. As the following lemma states, the modified conditions (with appropriate parameters) can be deduced from VM and EM.

**Lemma 3.5** (i) *For any $\varepsilon < \frac{1}{2}$, if $\mu = \mu_T^\tau$ satisfies $\mathrm{VM}(\ell,\varepsilon)$ then for every $x \in T$, any $\eta \in \Omega_T^\tau$ and any function $f$ we have $\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \le \frac{2-\varepsilon'}{1-\varepsilon'} \cdot \mu_{T_x}^\eta[\mathrm{Var}_{B_{x,\ell}}(f)] + \frac{\varepsilon'}{1-\varepsilon'} \cdot \mu_{T_x}^\eta[\mathrm{Var}_{\widetilde{T_x}}(f)]$, with $\varepsilon' = 2\varepsilon$.*

(ii) *For any $\varepsilon < p_{\min}^2$, if $\mu = \mu_T^\tau$ satisfies $\mathrm{EM}(\ell,\varepsilon)$ then for every $x \in T$, any $\eta \in \Omega_T^\tau$ and any function $f \ge 0$ we have $\mathrm{Ent}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \le \frac{1}{1-\varepsilon'} \cdot \mu_{T_x}^\eta[\mathrm{Ent}_{B_{x,\ell}}(f)] + \frac{\varepsilon'}{1-\varepsilon'} \cdot \mu_{T_x}^\eta[\mathrm{Ent}_{\widetilde{T_x}}(f)]$, with $\varepsilon' = \frac{\sqrt{\varepsilon}}{p_{\min}}$.*

**Remark:** We note that with extra work, part (ii) of Lemma 3.5 can be improved to hold with $\varepsilon' = c(p_{\min})\varepsilon$. We give the weaker bound because it is simpler to prove while still enough for our applications.

Similar statements to those in Lemma 3.5 appeared in [4]. We defer our proof to Section 7.

We can now prove Theorems 3.2 and 3.4 by working with the modified spatial mixing conditions of Lemma 3.5.

**Proof of Theorems 3.2 and 3.4:** Here we only prove the forward direction of both theorems. The reverse direction of Theorem 3.2 was proved in [3], as already mentioned above. The proof of the reverse direction of Theorem 3.4 is deferred to Section 7 because it uses machinery developed later in the paper.

The main step in the proof of the forward direction is to show the following claim:

**Claim 3.6** *If for every $x \in T$, any $\eta \in \Omega_T^\tau$ and any function $f$,*

$$\mathrm{Var}_{T_x}^\eta[\mu_{\widetilde{T_x}}(f)] \;\le\; c\cdot\mu_{T_x}^\eta[\mathrm{Var}_{B_{x,\ell}}(f)] \;+\; \left(\frac{1-\delta}{\ell}\right)\cdot\mu_{T_x}^\eta[\mathrm{Var}_{\widetilde{T_x}}(f)],$$

*then $\mathrm{Var}(f) \le \frac{c}{\delta}\cdot\mathcal{D}_\ell(f)$ for all $f$. The same implication holds when $\mathrm{Var}$ is replaced by $\mathrm{Ent}$, $\mathcal{D}_\ell$ is replaced by $\mathcal{E}_\ell$ and the function $f$ is restricted to be non-negative.*

Observe that the hypothesis of Theorem 3.2 together with part (i) of Lemma 3.5 establishes the hypothesis of Claim 3.6 with $c \le 3$, and similarly, the hypothesis of Theorem 3.4 together with part (ii) of Lemma 3.5 establishes the hypothesis of Claim 3.6 (after the necessary replacement of symbols) with $c \le 2$.

It therefore suffices to prove Claim 3.6. We prove only the formulation with $\mathrm{Var}$ and $\mathcal{D}_\ell$ since the proof for the formulation with $\mathrm{Ent}$ and $\mathcal{E}_\ell$ is identical once we make the same replacements in the text of the proof. As will be clear below, the proof uses only properties which are common to both $\mathrm{Var}$ and $\mathrm{Ent}$.

Consider an arbitrary function $f : \Omega \to \mathbb{R}$. Our first goal is to relate $\mathrm{Var}(f)$ to the projections $\mathrm{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(f)]$ for $x \in T$, so that we can apply the spatial mixing condition of the hypothesis. Recall that $T$ has $m+1$ levels, and define the increasing sequence $\emptyset = F_0 \subset F_1 \subset \ldots \subset F_{m+1} = T$, where $F_i$ consists of all sites in the lowest $i$ levels of $T$. Thus $F_i$ is a forest of height $i-1$. Using (2) recursively, and the facts that $\mu_{F_{i+1}}(\mu_{F_i}(f)) = \mu_{F_{i+1}}(f)$ and $\mu_{F_0}(f) = f$, we obtain

$$\begin{aligned}
\mathrm{Var}(f) &= \mu[\mathrm{Var}_{F_1}(f)] + \mathrm{Var}[\mu_{F_1}(f)] \\
&= \mu[\mathrm{Var}_{F_1}(f)] + \mu[\mathrm{Var}_{F_2}(\mu_{F_1}(f))] + \mathrm{Var}[\mu_{F_2}(\mu_{F_1}(f))] \\
&\ \ \vdots \\
&= \sum_{i=1}^{m+1} \mu[\mathrm{Var}_{F_i}(\mu_{F_{i-1}}(f))].
\end{aligned}$$

Now a fundamental property of nearest-neighbor interaction models on a tree is that, given the configuration on $T \setminus F_i$, the Gibbs distribution on $F_i$ becomes a product of the marginals on the subtrees rooted at the sites $x \in F_i \setminus F_{i-1}$. Using inequality (3) for the variance of a product measure, we therefore have that

$$\mathrm{Var}(f) \le \sum_{i=1}^{m+1} \sum_{x \in F_i \setminus F_{i-1}} \mu[\mathrm{Var}_{T_x}(\mu_{F_{i-1}}(f))] \le \sum_{x \in T} \mu[\mathrm{Var}_{T_x}(\mu_{\widetilde{T_x}}(f))], \tag{11}$$

where in the second inequality we used the convexity of the variance as in (4).

Notice that so far we have not used the spatial mixing condition in the hypothesis of Claim 3.6, but only a natural martingale structure induced by the tree. Let us denote the final sum in (11) by $\mathrm{Pvar}(f)$. In order to bound $c_{\mathrm{gap}}$, we need to compare the projection terms $\mathrm{Var}_{T_x}(\mu_{\widetilde{T_x}}(f))$ in $\mathrm{Pvar}(f)$ with the local conditional variance terms in $\mathcal{D}_\ell(f)$. For example, notice that if $\mu$ were the product of its single-site marginals then $\mathrm{Var}_{T_x}(\mu_{\widetilde{T_x}}(f)) \le \mu_{T_x}[\mathrm{Var}_x(f)]$ and $c_{\mathrm{gap}} = 1$. However, in general the variance of the projection on $x$ may also involve terms which depend on other sites, and may lead to a factor that grows with the size of $T_x$. We will use the spatial mixing condition in order to preclude the latter possibility. Specifically, we show that if for every $x \in T$, any $\eta \in \Omega^\tau_T$ and any function $g$, $\mathrm{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(g)] \;\le\; c \cdot \mu^\eta_{T_x}[\mathrm{Var}_{B_{x,\ell}}(g)] + \varepsilon \cdot \mu^\eta_{T_x}[\mathrm{Var}_{\widetilde{T_x}}(g)]$ then for every $x \in T$ and $\eta \in \Omega$,

$$\mathrm{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(f)] \;\le\; c \cdot \mu^\eta_{T_x}[\mathrm{Var}_{B_x}(f)] + \varepsilon \cdot \sum_{y \in B_x \cup \widetilde{\partial} B_x, y \ne x} \mu^\eta_{T_x}[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))], \tag{12}$$

where we have abbreviated $B_{x,\ell}$ to $B_x$ and $\widetilde{\partial} B_x$ stands for the boundary of $B_x$ excluding the parent of $x$, i.e., the bottom boundary of $B_x$. Notice that the last term in (12) is relevant only when $x$ is at distance at least $\ell$ from the bottom of $T$. When $x$ belongs to one of the $\ell$ lowest levels of $T$ then $T_x = B_x$, and thus trivially $\mathrm{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(f)] \le \mu^\eta_{T_x}[\mathrm{Var}_{B_x}(f)]$.

Let us assume (12) for now and conclude the proof of the theorem. Applying (12) for every $x$ and $\eta$, and using the hypothesis that $\varepsilon = \frac{1-\delta}{\ell}$ and the fact that each site appears in at most $\ell$ blocks,

we get

$$\begin{aligned}\mathrm{Pvar}(f) &\le c\cdot\mathcal{D}_\ell(f)+\varepsilon\cdot\sum_{x\in T}\sum_{y\in B_x\cup\widetilde{\partial}B_x, y\ne x}\mu[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))]\\ &\le c\cdot\mathcal{D}_\ell(f)+\varepsilon\ell\cdot\sum_{y\in T}\mu[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))]\\ &= c\cdot\mathcal{D}_\ell(f)+(1-\delta)\mathrm{Pvar}(f),\end{aligned}$$

and hence

$$\mathrm{Var}(f)\ \le\ \mathrm{Pvar}(f)\ \le\ \frac{c}{\delta}\cdot\mathcal{D}_\ell(f),$$

proving Claim 3.6. We now return to proving (12).

Let $g=\mu_{T_x\setminus(B_x\cup\widetilde{\partial}B_x)}(f)$. Once we notice that $\mu_{\widetilde{T_x}}(f)=\mu_{\widetilde{T_x}}(g)$, we can use the spatial mixing assumption that precedes (12) to deduce

$$\begin{aligned}\mathrm{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(f)] &\le c\cdot\mu^\eta_{T_x}[\mathrm{Var}_{B_x}(g)]+\varepsilon\cdot\mu^\eta_{T_x}[\mathrm{Var}_{\widetilde{T_x}}(g)]\\ &\le c\cdot\mu^\eta_{T_x}[\mathrm{Var}_{B_x}(f)]+\varepsilon\cdot\mu^\eta_{T_x}[\mathrm{Var}_{\widetilde{T_x}}(g)],\end{aligned}$$

where we used (4) for the second inequality. We will be done once we show that

$$\mu^\eta_{T_x}[\mathrm{Var}_{\widetilde{T_x}}(g)]\ \le \sum_{y\in B_x\cup\widetilde{\partial}B_x, y\ne x}\mu^\eta_{T_x}[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))]. \tag{13}$$

But (13) follows from a similar argument to that used earlier to show $\mathrm{Var}(f)\le\mathrm{Pvar}(f)$, starting from the fact that $g=\mu_{F'_k}(f)$, where the forests $F'_i$ are defined analogously to the $F_i$ earlier but restricted to the subtree $T_x$, and $k=\mathrm{height}(x)-\ell$. We omit the details.

This concludes the proof of Claim 3.6, and thus of Theorems 3.2 and 3.4. □

# 4 Verifying spatial mixing for the spectral gap

In this section, we will prove that the spectral gap of the Glauber dynamics is bounded in all of the situations covered by Theorem 1.1 in the Introduction.

In light of Theorem 3.2, to bound the spectral gap it suffices to verify the Variance Mixing condition $\mathrm{VM}(\ell,\varepsilon)$ with $\varepsilon=(1-\delta)/2(\ell+1-\delta)$, for some constants $\ell,\delta>0$ independent of the size of $T$. In fact, we will show it with the asymptotically tighter value $\varepsilon=c\exp(-\vartheta\ell)$:

**Theorem 4.1** *In both of the following situations, there exists a positive constant $\vartheta$ (depending only on $b,\beta$ and $h$) such that, for all $T$, the Gibbs distribution $\mu=\mu^\tau_T$ satisfies $\mathrm{VM}(\ell,e^{-\vartheta\ell})$ for all $\ell$:*

(i) *$\tau$ is arbitrary, and either $\beta<\beta_1$ (with $h$ arbitrary), or $|h|>h_c(\beta)$ (with $\beta$ arbitrary);*

(ii) *$\tau$ is the $(+)$-boundary condition, and $\beta,h$ are arbitrary.*

*As a corollary, in both situations $c_{\mathrm{gap}}(\mu)=\Omega(1)$.*

**Remark:** The validity of VM, i.e, the decay of point-to-set correlations, is of interest independently of its implication for the spectral gap (an implication which is new to this paper): e.g., it is closely related to the purity of the infinite volume Gibbs measure and to bit reconstruction problems on trees [13]. In the special case of a free boundary and $h=0$, part (i) of Theorem 4.1 was first proved in [6] via a lengthy calculation, which was considerably simplified in [19]. It was later reproved in [3] (for arbitrary boundary conditions) as a consequence of the fact that the spectral gap is bounded in this situation. An extension to general trees can be found in [13] and [20]. Our motivation for presenting another proof of part (i) (in addition to handling general fields $h$) is the simplicity of our argument compared with previous ones. As far as part (ii) is concerned, we are unaware of any previous results for the case of the $(+)$-boundary other than the fact that $\mathrm{VM}(\ell,\varepsilon(\ell))$ must hold with $\lim_{\ell\to\infty}\varepsilon(\ell)=0$ because the $(+)$-phase is pure (see, e.g., [15]).

The rest of this section is divided into two parts. First, we develop a general framework based on coupling in order to establish the exponential decay of point-to-set correlations. This framework identifies two key quantities, $\kappa$ and $\gamma$, and states that when their product is small enough then VM holds. Then, in the second part, we go back to proving Theorem 4.1 by calculating $\kappa$ and $\gamma$ for each of the above two regimes separately.

## 4.1 A coupling argument for decay of point-to-set correlations

In this section we develop a coupling framework that enables us to verify the exponential decay of point-to-set correlations from a simple calculation involving single-spin distributions.

First we need some additional notation. When $x$ is not the root of $T$, let $\mu^+_{T_x}$ (respectively, $\mu^-_{T_x}$) denote the Gibbs distribution in which the parent of $x$ has its spin fixed to $(+)$ (respectively, $(-)$) and the configuration on the bottom boundary of $T_x$ is specified by $\tau$ (the global boundary condition on $T$) [¶]. For two distributions $\mu_1$ and $\mu_2$, we denote by $\|\mu_1 - \mu_2\|_x$ the variation distance between the projections of $\mu_1$ and $\mu_2$ onto the spin at $x$. (Since the Ising model has only two spin values, $\|\mu_1 - \mu_2\|_x = |\mu_1(\sigma_x = +) - \mu_2(\sigma_x = +)|$.) Recall also that $\eta^y$ denotes the configuration $\eta$ with the spin at site $y$ flipped.

We now identify two constants that are crucial for our coupling argument:

**Definition 4.2** *For a sequence of Gibbs distributions $\{\mu^\tau_T\}$ corresponding to a fixed boundary condition $\tau$, define $\kappa \equiv \kappa(\{\mu^\tau_T\})$ and $\gamma \equiv \gamma(\{\mu^\tau_T\})$ by*

(i) $\kappa = \sup_T \max_z \|\mu^+_{T_z} - \mu^-_{T_z}\|_z$;

(ii) $\gamma = \sup_T \max \|\mu^\eta_A - \mu^{\eta^y}_A\|_z$, *where the maximum is taken over all subsets $A \subseteq T$, all boundary configurations $\eta$, all sites $y$ on the boundary of $A$ and all neighbors $z \in A$ of $y$.*

Note that $\kappa$ is the same as $\gamma$, except that the maximization is restricted to $A = T_z$ and the boundary vertex $y$ being the parent of $z$; hence always $\kappa \le \gamma$. Since $\kappa$ involves Gibbs distributions only on maximal subtrees $T_z$, it may depend on the boundary condition $\tau$ at the bottom of the tree. By contrast, $\gamma$ bounds the worst-case probability of disagreement for an *arbitrary* subset $A$ and arbitrary boundary configuration around $A$, and hence depends only on $(\beta, h)$ and not on $\tau$. It is the dependence of $\kappa$ on $\tau$ that opens up the possibility of an analysis that is specific to the boundary condition. For example, at very low temperature and with no external field, $\kappa$ is close to $1$ in the free boundary case, while it is close to zero in the $(+)$-boundary case.

In our arguments $\kappa$ will be used to bound the probability of a disagreement percolating one level *down* the tree, namely, when we fix a disagreement at $x$ and couple the two resulting marginals on a child $z$ of $x$. On the other hand, $\gamma$ will be used in order to bound the probability of a disagreement percolating one level *up* the tree, namely, when we fix a single disagreement on the bottom boundary of a block, say at $y$ (with the rest of the boundary configuration being arbitrary), and couple the marginals on the parent of $y$.

The novelty of our argument for establishing VM comes from the fact that we identify *two separate* constants $\kappa$ and $\gamma$, and consider their product, rather than working with $\kappa$ alone:

**Theorem 4.3** *Any Gibbs distribution $\mu = \mu^\tau_T$ satisfies $\mathrm{VM}(\ell, (\gamma\kappa b)^\ell)$ for all $\ell$, where $\kappa$ and $\gamma$ are the constants associated with the sequence $\{\mu^\tau_T\}$ as specified in Definition 4.2. In particular, if $\gamma\kappa b < 1$ then there exists a constant $\vartheta > 0$ such that, for every $T$, the measure $\mu = \mu^\tau_T$ satisfies $\mathrm{VM}(\ell, e^{-\vartheta\ell})$ for all $\ell$, and hence $c_{\mathrm{gap}}(\mu) = \Omega(1)$.*

[¶]Notice that we do not specify the rest of the configuration outside $T_x$ since it has no influence on the distribution inside $T_x$ once the spin at the parent of $x$ is fixed. However, since our distributions are defined over the whole configuration space, in the discussion below when the configuration outside $T_x$ is relevant it will be understood from the context.

**Proof:** Fix arbitrary $T$, $x \in T$, $\eta \in \Omega_T^\tau$. We need to show that for every function $f$ that does not depend on $B_{x,\ell}$, $\mathrm{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(f)] \le \varepsilon \cdot \mathrm{Var}^\eta_{T_x}(f)$ with $\varepsilon = (\kappa\gamma b)^\ell$, i.e., projecting $f$ onto the root (of $T_x$) causes the variance to shrink by a factor $\varepsilon$. As is well known, it is enough to establish a dual contraction, i.e., to consider an arbitrary function that depends only on the spin at the root and show that, when projecting onto levels $\ell$ and below, the variance shrinks by a factor $\varepsilon$. Formally, it is enough to show that for every function $g$ that does not depend on $\widetilde{T_x}$[‖] we have

$$\mathrm{Var}^\eta_{T_x}[\mu_{B_{x,\ell}}(g)] \;\le\; \varepsilon \cdot \mathrm{Var}^\eta_{T_x}(g). \tag{14}$$

This is because for a function $f$ that does not depend on $B_{x,\ell}$, the variance of the projection can be written as

$$\mathrm{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(f)] \;=\; \mathrm{Cov}^\eta_{T_x}(f, \mu_{\widetilde{T_x}}(f)) \;=\; \mathrm{Cov}^\eta_{T_x}(f, \mu_{B_{x,\ell}}(\mu_{\widetilde{T_x}}(f))) \;\le\; \sqrt{\mathrm{Var}^\eta_{T_x}(f) \cdot \mathrm{Var}^\eta_{T_x}[\mu_{B_{x,\ell}}(\mu_{\widetilde{T_x}}(f)]}\ ,$$

where $\mathrm{Cov}^\eta_A(f,f')$ denotes the covariance $\mu^\eta_A(ff') - \mu^\eta_A(f)\mu^\eta_A(f')$ and the last inequality is an application of Cauchy-Schwartz. We then have

$$\mathrm{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(f)] \;\le\; \mathrm{Var}^\eta_{T_x}(f) \cdot \frac{\mathrm{Var}^\eta_{T_x}[\mu_{B_{x,\ell}}(\mu_{\widetilde{T_x}}(f))]}{\mathrm{Var}^\eta_{T_x}[\mu_{\widetilde{T_x}}(f)]}\ .$$

If we assume (14) then the expression on the r.h.s. is bounded by $\varepsilon \cdot \mathrm{Var}^\eta_{T_x}(f)$ since $g = \mu_{\widetilde{T_x}}(f)$ does not depend on $\widetilde{T_x}$.

We therefore proceed with the proof of (14), which goes via a coupling argument. A *coupling* of two distributions $\mu_1, \mu_2$ on $\Omega$ is any joint distribution $\nu$ on $\Omega^2$ whose marginals are $\mu_1$ and $\mu_2$ respectively. For two configurations $\sigma, \sigma' \in \Omega$, let $|\sigma - \sigma'|_{x,\ell}$ denote the *Hamming distance* between the restrictions of $\sigma$ and $\sigma'$ to $\widetilde{\partial} B_{x,\ell}$, i.e., the number of sites at distance $\ell$ below $x$ at which $\sigma$ and $\sigma'$ differ. Notice that $|\sigma - \sigma'|_{x,\ell}$ can be at most $b^\ell$, the number of sites on the $\ell$th level below $x$. Let $\mu^+_{\widetilde{T_x}}$ (respectively, $\mu^-_{\widetilde{T_x}}$) stand for the Gibbs distribution where the spin at $x$ is set to $(+)$ (respectively, $(-)$) and, as usual, the configuration on the bottom boundary of $\widetilde{T_x}$ is specified by $\tau$. Our goal will be to construct a coupling $\nu$ of $\mu^+_{\widetilde{T_x}}$ and $\mu^-_{\widetilde{T_x}}$ for which the expectation $\mathrm{E}_\nu|\sigma - \sigma'|_{x,\ell} \equiv \sum_{\sigma,\sigma'} \nu(\sigma,\sigma')|\sigma - \sigma'|_{x,\ell}$ is only $(\kappa b)^\ell$.

**Claim 4.4** *For every $x \in T$ and all $\ell$ the following hold*:

(i) *There is a coupling $\nu$ of $\mu^+_{\widetilde{T_x}}$ and $\mu^-_{\widetilde{T_x}}$ for which $\mathrm{E}_\nu|\sigma - \sigma'|_{x,\ell} \le (\kappa b)^\ell$.*

(ii) *For any $\eta, \eta' \in \Omega$ that have the same spin value at the parent of $x$, $\|\mu^\eta_{B_{x,\ell}} - \mu^{\eta'}_{B_{x,\ell}}\|_x \le \gamma^\ell \cdot |\eta - \eta'|_{x,\ell}$.*

Let us assume Claim 4.4 for the moment and complete the proof of (14). Consider an arbitrary $g$ that does not depend on $\widetilde{T_x}$. Let $p = \mu^\eta_{T_x}(\sigma_x = +)$ and $q = 1 - p = \mu^\eta_{T_x}(\sigma_x = -)$. We also write $g^+$ for $g(\sigma)$, where $\sigma$ is any configuration that agrees with $\eta$ outside $T_x$ and such that $\sigma_x = +$. (This is well defined since $g$ does not depend on $\widetilde{T_x}$). We define $g^-$ similarly. Without loss of generality we may assume that in the coupling $\nu$ from Claim 4.4 both the coupled configurations agree with $\eta$

[‖] Effectively this means that, conditioned on the configuration outside $T_x$ being $\eta$, $g$ depends only on the spin at the root $x$.

outside $T_x$ with probability 1. We then have

$$
\begin{aligned}
\mathrm{Var}^{\eta}_{T_x}[\mu_{B_{x,\ell}}(g)] &= \mathrm{Cov}^{\eta}_{T_x}[g, \mu_{B_{x,\ell}}(g)] \\
&= \mathrm{Cov}^{\eta}_{T_x}[g, \mu_{\widetilde{T_x}}(\mu_{B_{x,\ell}}(g))] \\
&= pq(g^+ - g^-)[\mu^+_{\widetilde{T_x}}(\mu_{B_{x,\ell}}(g)) - \mu^-_{\widetilde{T_x}}(\mu_{B_{x,\ell}}(g))] \\
&= pq(g^+ - g^-)\sum_{\sigma,\sigma'} \nu(\sigma,\sigma')[\mu^{\sigma}_{B_{x,\ell}}(g) - \mu^{\sigma'}_{B_{x,\ell}}(g)] \\
&\le pq|g^+ - g^-|\sum_{\sigma,\sigma'} \nu(\sigma,\sigma')\|\mu^{\sigma}_{B_{x,\ell}} - \mu^{\sigma'}_{B_{x,\ell}}\|_x \cdot |g^+ - g^-| \qquad (15)\\
&\le pq(g^+ - g^-)^2\sum_{\sigma,\sigma'} \nu(\sigma,\sigma')|\sigma - \sigma'|_{x,\ell} \cdot \gamma^{\ell} \\
&= \gamma^{\ell} \cdot \mathrm{Var}^{\eta}_{T_x}(g) \cdot \mathrm{E}_{\nu}|\sigma - \sigma'|_{x,\ell} \\
&\le (\gamma\kappa b)^{\ell} \cdot \mathrm{Var}^{\eta}_{T_x}(g).
\end{aligned}
$$

In the sixth line here we have used part (ii) of Claim 4.4, and in the last line we have used part (i). This completes the proof of (14), and hence of Theorem 4.3. We thus go back and prove Claim 4.4.

The proof of Claim 4.4 makes use of a standard recursive coupling along paths in the tree (as in, e.g., [3]). We start with part (i), i.e., constructing a coupling $\nu$ of $\mu^+_{\widetilde{T_x}}$ and $\mu^-_{\widetilde{T_x}}$ with the required properties. Since the underlying graph is a tree, we can couple $\mu^+_{\widetilde{T_x}}$ and $\mu^-_{\widetilde{T_x}}$ recursively. This goes as follows. First, given the spin at $x$ the measures on $T_z$ (where $z$ ranges over the children of $x$) are all independent of each other, so we can couple the projections on the $T_z$'s independently. Then, we couple the two projections on $T_z$ by first coupling the spin at $z$ using the optimal coupling (the one that achieves the variation distance) of the marginal measures on the spin at $z$. Thus, the spins at $z$ disagree with probability at most $\kappa$. Once a coupled pair of spins at $z$ is chosen, we continue as follows: if the spins at $z$ agree then we can make the configurations in $\widetilde{T_z}$ equal with probability 1 (because the two boundary conditions are the same); if the spins at $z$ differ (i.e., one is $(+)$ and the other $(-)$) then we recursively couple $\mu^+_{\widetilde{T_z}}$ and $\mu^-_{\widetilde{T_z}}$. We let $\nu$ be the resulting coupling of $\mu^+_{\widetilde{T_x}}$ and $\mu^-_{\widetilde{T_x}}$, and notice that $\mathrm{E}_{\nu}|\sigma - \sigma'|_{x,l} \le (\kappa b)^{\ell}$ since for every site $y$ at distance $\ell$ below $x$ the probability that the two coupled spins at $y$ disagree is at most $\kappa^{\ell}$.

We go on to prove part (ii) of Claim 4.4. First, by writing a telescopic sum and applying the triangle inequality we get that

$$
\|\mu^{\eta}_{B_{x,\ell}} - \mu^{\eta'}_{B_{x,\ell}}\|_x \ \le\ \sum_{i=1}^{k} \|\mu^{\eta^{(i-1)}}_{B_{x,\ell}} - \mu^{\eta^{(i)}}_{B_{x,\ell}}\|_x \ \ ,
$$

where $k = |\eta - \eta'|_{x,\ell}$ and the sequence of configurations $\eta^{(i)}$ is a site-by-site interpolation of the differences between $\eta$ and $\eta'$ in $\widetilde{\partial} B_{x,\ell}$. (It suffices to interpolate only over the differences in $\widetilde{\partial} B_{x,\ell}$ since the measure $\mu^{\eta}_{B_{x,\ell}}$ depends only on the configuration in $\partial B_{x,\ell}$ and since $\eta$ and $\eta'$ agree on the parent of $x$.) It is now enough to show that $\|\mu^{\eta}_{B_{x,\ell}} - \mu^{\eta^w}_{B_{x,\ell}}\|_x \le \gamma^{\ell}$ for all $\eta$ and $w \in \widetilde{\partial} B_{x,\ell}$. This, however, follows by a coupling argument as before, where this time we couple recursively along the path from $w$ to $x$ (i.e., up the tree). Specifically, suppose by induction that in our coupling there is already a path of disagreement going from $w$ to $y$, where $y$ is some site on the path from $w$ to $x$. Let $z$ denote the parent of $y$. At the next step we choose a coupled pair of spins at $z$ from the two distributions $\mu^{\eta}_A$ and $\mu^{\eta^y}_A$ (using an optimal coupling for the projections onto the spin at $z$), where the subset $A$ is $B_{x,\ell}$ excluding the path from $w$ to $y$. The probability of disagreement at $z$ given the disagreement at $y$ is then bounded by $\gamma$, by definition. If the resulting spins at $z$ agree then the spins on the rest of the path are coupled to agree with certainty, while if there is a disagreement

at $z$ we continue recursively starting from the disagreement at $z$. We therefore conclude that the probability of disagreement at $x$ in the resulting coupling is $\gamma^\ell$, as required. $\square$

**Remark:** We emphasize that Theorem 4.3 is not specific to the Ising model and generalizes to arbitrary nearest-neighbor models on a tree. Although we used the fact that the Ising model has only two possible spin values, the proof can easily be generalized to more than two spin values at the cost of a factor $\frac{1}{p_{\min}}$ in front of $(\gamma\kappa b)^\ell$ in $\mathrm{VM}$, where $p_{\min}$ is the minimum probability of any spin value as defined just before Theorem 3.4. Thus, since Theorem 3.2 also applies to general nearest-neighbor spin systems on a tree, we conclude that the implication from $\gamma\kappa b < 1$ to a bounded $c_{\mathrm{gap}}(\mu)$ holds for any such system (with the definitions of $\kappa$ and $\gamma$ extended in the obvious way to systems with more than two spin values). The details can be found in the companion paper [31].

## 4.2 Proof of Theorem 4.1

In this section we go back to proving Theorem 4.1. Using Theorem 4.3, all we need to do for the given choices of the Ising model parameters is to bound $\kappa$ and $\gamma$ as in Definition 4.2 such that $\gamma\kappa b < 1$. In contrast to Sections 3 and 4.1, which apply to general nearest-neighbor spin systems on trees, here the calculations are specific to the Ising model.

For both $\kappa$ and $\gamma$, we need to bound a quantity of the form $\|\mu_A^\eta - \mu_A^{\eta^y}\|_z$, where $y \in \partial A$ and $z \in A$ is a neighbor of $y$. The key observation is that this quantity can be expressed very cleanly in terms of the "magnetization" at $z$, i.e., the ratio of probabilities of a $(-)$-spin and a $(+)$-spin at $z$. It will actually be convenient to work with the magnetization *without* the influence of the neighbor $y$: thus we let $\mu_A^{\eta,y=*}$ denote the Gibbs distribution with boundary condition $\eta$, except that the spin at $y$ is free (or equivalently, the edge connecting $z$ to $y$ is erased). We then have:

**Proposition 4.5** *For any subset $A \subseteq T$, any boundary configuration $\eta$, any site $y \in \partial A$ and any neighbor $z \in A$ of $y$, we have*

$$\|\mu_A^\eta - \mu_A^{\eta^y}\|_z = K_\beta(R),$$

*where $R = \frac{\mu_A^{\eta,y=*}(\sigma_z=-)}{\mu_A^{\eta,y=*}(\sigma_z=+)}$ and the function $K_\beta$ is defined by*

$$K_\beta(a) = \frac{1}{e^{-2\beta}a+1} - \frac{1}{e^{2\beta}a+1}.$$

**Proof:** First, w.l.o.g. we may assume that the edge between $y$ and $z$ is the only one connecting $y$ to $A$; this is because a tree has no cycles, so once the spin at $y$ is fixed $A$ decomposes into disjoint components that are independent. We also assume w.l.o.g. that the spin at $y$ is $(+)$ in $\eta$, and we abbreviate $\mu_A^\eta$ and $\mu_A^{\eta^y}$ to $\mu_A^+$ and $\mu_A^-$ respectively, and also $\mu_A^{\eta,y=*}$ to $\mu_A^*$. Thus $\|\mu_A^\eta - \mu_A^{\eta^y}\|_z = |\mu_A^+(\sigma_z=+) - \mu_A^-(\sigma_z=+)|$, and $R = \frac{\mu_A^*(\sigma_z=-)}{\mu_A^*(\sigma_z=+)}$. We write $R^+$ for $\frac{\mu_A^+(\sigma_z=-)}{\mu_A^+(\sigma_z=+)}$ and $R^-$ for $\frac{\mu_A^-(\sigma_z=-)}{\mu_A^-(\sigma_z=+)}$. Since the only influence of $y$ on $A$ is through $z$, we have $R^+ = e^{-2\beta}R$ and $R^- = e^{2\beta}R$. The proposition now follows once we notice that, by definition of $R^+$ and $R^-$, $\mu_A^+(\sigma_z=+) = \frac{1}{R^++1}$ and $\mu_A^-(\sigma_z=+) = \frac{1}{R^-+1}$. $\square$

Now it is easy to check that $K_\beta(a)$ is an increasing function in the interval $[0,1]$, decreasing in the interval $[1,\infty]$, and is maximized at $a=1$. Therefore, we can always bound $\kappa$ and $\gamma$ from above by $K_\beta(1) = \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$. Indeed, for $\gamma$ we must make do with this crude bound because it has to hold for any boundary configuration $\eta$ and we cannot hope to gain by controlling the magnetization $R$. However, as we shall see, for $\kappa$ we can do better in some cases by computing the magnetization at the root; when this differs from $1$ we get a better bound than $K_\beta(1)$.

We are now ready to proceed to the proof of Theorem 4.1:

**(i) Arbitrary boundary conditions**

Here, the boundary condition $\tau$ is arbitrary and we first consider the (easy) case when $\beta < \beta_0$ or $|h| > h_c(\beta)$ (i.e., $h$ is super-critical). In this case we do not need to resort to the calculation of $\kappa$ and $\gamma$. As discussed in the Introduction, in this regime there is a unique infinite volume Gibbs measure, so certainly the variation distance at the root $\max_{\eta,\eta'} \|\mu^{\eta}_{B_{x,\ell}} - \mu^{\eta'}_{B_{x,\ell}}\|_x$ goes to zero as $\ell$ increases. In fact, it is not too difficult to see that in the above regime this variation distance goes to zero exponentially fast, which directly implies the desired exponential decay of correlations (VM) by plugging the bound on the variation distance into expression (15) in the proof of Theorem 4.3.

We go on to consider the more interesting regime when $\beta_0 \le \beta < \beta_1$ (i.e., intermediate temperatures) and the external field $h$ is arbitrary. Here we use the fact that $\kappa \le \gamma \le K_\beta(1)$. We then certainly have $\gamma\kappa b < 1$ whenever $K_\beta(1) = \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} < \sqrt{\frac{1}{b}}$, i.e., whenever $e^{-2\beta} > \frac{\sqrt{b}-1}{\sqrt{b}+1}$. From the definition of $\beta_1$ (see Section 1.1), this corresponds precisely to $\beta < \beta_1$. (Observe how this non-trivial result drops out immediately from our machinery, as expressed in the condition $\gamma^2 < \frac{1}{b}$.)

This completes the verification of Theorem 4.1 part (i).

**(ii) $(+)$-boundary condition**

We now assume that $\tau$ is the all-$(+)$ configuration and consider arbitrary $\beta$ and $h$. For convenience, we assume $h \ge -h_c(\beta)$ since the case $|h| > h_c(\beta)$ was covered in part (i) for all boundary conditions $\tau$. The important property of the regime $h \ge -h_c(\beta)$ is that, for the $(+)$-boundary, the spin at the root is at least as likely to be $(+)$ as it is to be $(-)$. We will show that $\gamma\kappa b < 1$ throughout this regime. Recall that we already showed that $\gamma \le K_\beta(1) < 1$ for all finite $\beta$. It is therefore enough to show that $\kappa \le \frac{1}{b}$.

To calculate $\kappa$, we need to bound the variation distance $\|\mu^+_{T_z} - \mu^-_{T_z}\|_z$, which by Proposition 4.5 is equal to $K_\beta(R_z)$, where $R_z = \frac{\mu^*_{T_z}(\sigma_z = -)}{\mu^*_{T_z}(\sigma_z = +)}$ and $\mu^*_{T_z}$ is the Gibbs distribution over the subtree $T_z$ when it is disconnected from the rest of $T$ and the spins on its bottom boundary agree with $\tau$. We thus have $\kappa = \sup_T \max_{z \in T} K_\beta(R_z)$.

The final ingredient we need is a recursive computation of the magnetization $R_z$, the details of which (up to change of variables) can be found in [2] or [5]. Let $y \prec z$ denote that $y$ is a child of $z$. A simple direct calculation gives that $R_z = e^{-2\beta h} \prod_{y \prec z} F(R_y)$, where $F(a) \equiv F_\beta(a) = \frac{a + e^{-2\beta}}{e^{-2\beta} a + 1}$. In particular, if $z$ is any site on the bottom-most level of $T$, then since the spins of the children of $z$ are all set deterministically to $(+)$, we get that $R_z = e^{-2\beta h}[F(0)]^b$. We thus define

$$J(a) \equiv J_{\beta,h}(a) = e^{-2\beta h}[F(a)]^b \tag{16}$$

and observe that, for any $z \in T$, $R_z = J^{(\ell)}(0)$, where $J^{(\ell)}$ stands for the $\ell$-fold composition of $J$, and $\ell$ is the distance of $z$ from the bottom boundary of $T$.

We now describe some properties of $J$ that we use (refer to Fig. 2): $J$ is continuous and increasing on $[0, \infty)$, with $J(0) = e^{-2\beta(h+b)} > 0$ and $\sup_a J(a) = e^{-2\beta(h-b)} < \infty$. This immediately implies that $J$ has at least one fixed point in $[0, \infty)$; we denote by $a_0$ the least fixed point. Since $a_0$ is the least fixed point and $J(0) > 0$ then clearly $J'(a_0) \le 1$, where $J'(a) \equiv \frac{\partial J(a)}{\partial a}$ is the derivative of $J$. We also note that $a_0 \le 1$ when $h \ge -h_c(\beta)$, which corresponds to the fact that for the $(+)$-boundary and the above regime of $h$, the spin at the root is at least as likely to be $(+)$ as $(-)$.

Now, since $J$ is monotonically increasing and $a_0$ is the least fixed point of $J$, clearly $J^{(\ell)}(0)$ converges to $a_0$ from below, i.e., $R_z \le a_0$ for every $z \in T$. Thus, since $a_0 \le 1$ for $h \ge -h_c(\beta)$, and the function $K_\beta(a)$ is monotonically increasing in the interval $[0,1]$, $K_\beta(R_z) \le K_\beta(a_0)$ for every $z \in T$.

What remains to be shown is that $K_\beta(a_0) \le \frac{1}{b}$. This follows from the fact that $J'(a_0) \le 1$, together with the following lemma:

**Lemma 4.6** *Let $a_0$ be any fixed point of $J$. Then $K_\beta(a_0) = \frac{1}{b} \cdot J'(a_0)$.*

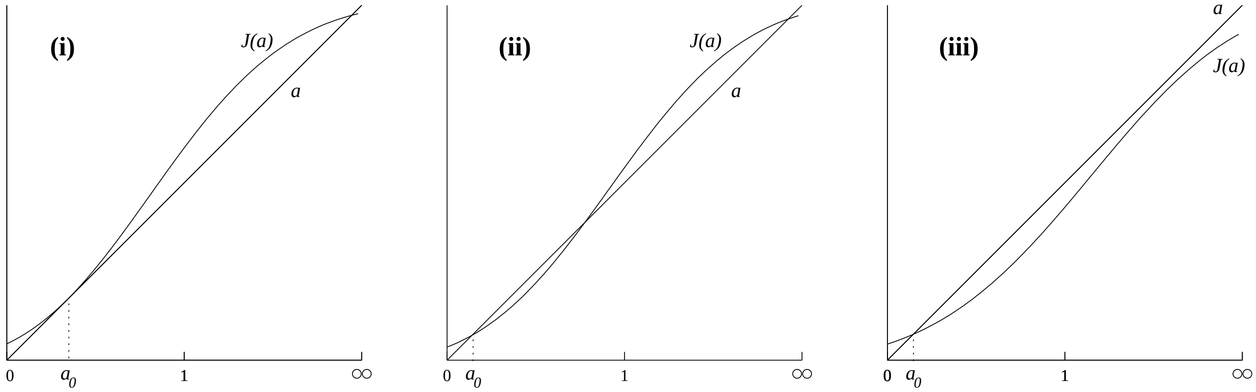


Figure 2: Curve of the function $J(a)$, used in the proof of Theorem 4.1, for $\beta > \beta_0$ and various values of the external field $h$. (i) $h = -h_c(\beta)$; (ii) $h_c(\beta) > h > -h_c(\beta)$; (iii) $h > h_c(\beta)$. The point $a_0$ is the smallest fixed point of $J$.

**Proof:** From the definitions of $J$ and $F$ we have:

$$\begin{aligned} J'(a_0) &= e^{-2\beta h} \cdot b \cdot [F(a_0)]^{b-1} F'(a_0) \\ &= b \cdot J(a_0) \cdot \frac{F'(a_0)}{F(a_0)} \\ &= b \cdot a_0 \cdot \frac{F'(a_0)}{F(a_0)} \\ &= b \cdot a_0 \cdot \Big[\frac{1 - e^{-4\beta}}{(a_0 + e^{-2\beta})(e^{-2\beta} a_0 + 1)}\Big] \\ &= b \cdot K_\beta(a_0). \qquad \Box \end{aligned}$$

This completes the verification of Theorem 4.1 part (ii).

# 5 Verifying spatial mixing for log-Sobolev

In this section we will prove a uniform lower bound (independent of $n$) on the logarithmic Sobolev constant $c_{\rm sob}(\mu)$ in all the situations covered by Theorem 1.2 in the Introduction.

In light of Theorem 3.4, to show $c_{\rm sob} = \Omega(1)$ we need only prove the validity of the Entropy Mixing condition $\mathrm{EM}(\ell, [(1-\delta)p_{\min}/2(\ell+1-\delta)]^2)$ for some constants $\ell$ and $\delta$ independent of the size of $T$. In order to establish EM in the situations covered by Theorem 1.2, we extend the coupling framework developed in Section 4.1 so that it can be used to establish EM. As before, we will use a condition on the constants $\kappa$ and $\gamma$, which were defined in Section 4.1. In fact, the condition on $\kappa$ and $\gamma$ for establishing EM is practically the same as the one that was used to establish VM, which immediately transfers our $\Omega(1)$ bound on $c_{\rm gap}$ for the relevant parameters to an $\Omega(1)$ bound on $c_{\rm sob}$ for the same choice of parameters. The main result of this section is the following relationship between $(\kappa, \gamma)$ and EM.

**Theorem 5.1** *Any Gibbs distribution $\mu = \mu_T^\tau$ satisfies $\mathrm{EM}(\ell, c(\gamma\alpha)^{\ell/5})$ for all $\ell$, where $\alpha = \max\{\kappa b, 1\}$, $\kappa$ and $\gamma$ are the constants associated with the sequence $\{\mu_T^\tau\}$ as specified in Definition 4.2, and $c$ is a constant that depends only on $(b, \beta, h)$. In particular, if $\max\{\gamma\kappa b, \gamma\} < 1$ then there exists a constant $\vartheta$ such that, for every $T$, the measure $\mu = \mu_T^\tau$ satisfies $\mathrm{EM}(\ell, ce^{-\vartheta\ell})$ for all $\ell$, and hence $c_{\rm sob}(\mu) = \Omega(1)$.*

**Remark:** We should note that the above theorem, like its counterpart for the spectral gap, holds for any spin system on a tree (with the definitions of $\kappa$ and $\gamma$ generalized appropriately). See the companion paper [31] for details.

Since in Section 4.2 we have already calculated $\kappa$ and $\gamma$ for the regimes of interest and shown that in both cases $\max\{\gamma\kappa b, \gamma\} < 1$, we have:

**Corollary 5.2** *In both of the following situations, $c_{\rm sob}(\mu) = \Omega(1)$:*

(i) *$\tau$ is arbitrary, and either $\beta < \beta_1$ (with $h$ arbitrary), or $|h| > h_c(\beta)$ (with $\beta$ arbitrary);*

(ii) *$\tau$ is the $(+)$-boundary condition and $\beta, h$ are arbitrary.*

This completes the proof of our second main result, Theorem 1.2 stated in the Introduction.

The first step in proving Theorem 5.1 is a reduction of EM to a certain *strong concentration* property of $\mu$, the Gibbs measure under consideration. We believe that this concentration property, as well as its connection to EM, may be of independent interest. The statement of this property and the reduction of EM to it is the content of Section 5.1. Then, in Section 5.2, we complete the proof of Theorem 5.1 by relating the strong concentration property to $\kappa$ and $\gamma$.

It is worth mentioning that we are also able to establish a general (but cruder) bound on $c_{\rm sob}$ as a function of $c_{\rm gap}$. Specifically, we can show that $c_{\rm sob} = \Omega(1/\log n) \times c_{\rm gap}$. Although we do not need this bound in this paper, we present it in Section 5.3 for future reference since its proof is simple and short.

## 5.1 Establishing EM via a strong concentration property.

In this subsection we reduce EM to a certain strong concentration property of $\mu$. In the next subsection, we will then establish this strong concentration property as a function of $\kappa$ and $\gamma$ in order to prove Theorem 5.1. For simplicity and without loss of generality, we will analyze the entropy mixing condition only for $T_x = T$ (the whole tree), with root $r$.

Let $\mu_{\widetilde{T}}^+$ and $\mu_{\widetilde{T}}^-$ denote the Gibbs distributions on $\widetilde{T}$ with the spin at the root $r$ set to $(+)$ and $(-)$ respectively (the boundary condition on the leaves of $T$ being specified by $\tau$). Define

$$g_+(\sigma) = \frac{\mu_{\widetilde{T}}^+(\sigma)}{\mu(\sigma)} = \begin{cases} 1/p & \text{if } \sigma_r = (+), \\ 0 & \text{otherwise,} \end{cases}$$

where $p = \mu(\sigma_r = +)$. The key quantity we will work with in the sequel is the following:

$$g_+^{(\ell)} = \mu_{B_{r,\ell}}(g_+).$$

Note that $g_+^{(\ell)}(\sigma)$ depends only on the spins in $\partial B_{r,\ell}$. Indeed, let $\sigma_{r,\ell}$ stand for the restriction of $\sigma$ to $\partial B_{r,\ell}$, i.e., to the sites at distance $\ell$ below $r$. It is easy to verify that $g_+^{(\ell)}(\sigma)$ is equal to $\frac{\mu_{\widetilde{T}}^+(\sigma_{r,\ell})}{\mu(\sigma_{r,\ell})}$. Thus, for a given configuration $\sigma$, $g_+^{(\ell)}(\sigma)$ is the ratio of the probabilities of seeing the spins of $\sigma$ at level $\ell$ below the root $r$ when the spin at $r$ is $(+)$ and when there is no condition on the spin at $r$, respectively. We define $g_-$ and $g_-^{(\ell)}$ in an analogous way.

The role played by the functions $g_+^{(\ell)}$ and $g_-^{(\ell)}$ is embodied in the following theorem, which says that if these functions are sufficiently tightly concentrated around their common mean value of 1 then the entropy mixing condition EM holds.

**Theorem 5.3** *There exists a constant $c$ (depending only on $b$, $\beta$ and $h$) such that, for any $\delta \geq 0$, if*

$$\mu\big[|g_s^{(\ell)} - 1| > \delta\big] \;\leq\; e^{-2/\delta} \tag{17}$$

*for $s \in \{+,-\}$, then we have $\mathrm{Ent}\big[\mu_{\widetilde{T}}(f)\big] \leq c\delta\,\mathrm{Ent}(f)$ for any non-negative function $f$ that does not depend on $B_{r,\ell}$; in particular, $\mathrm{EM}(\ell, c\delta)$ holds.*

**Proof:** Fix $\ell < m$ and a non-negative function $f$ that does not depend on the spins inside the block $B_{r,\ell}$. Since $\mathrm{Ent}(f') \le \mathrm{Var}(f')/\mu(f')$ for every non-negative function $f'$ (see, e.g., [36]) then

$$\begin{aligned}
\mathrm{Ent}\big[\mu_{\widetilde{T}}(f)\big] \;\le\; \frac{\mathrm{Var}\big[\mu_{\widetilde{T}}(f)\big]}{\mu\big[\mu_{\widetilde{T}}(f)\big]} \;&=\; \frac{1}{\mu(f)}\cdot\Big[p\big[\mu^{+}_{\widetilde{T}}(f)-\mu(f)\big]^2+(1-p)\big[\mu^{-}_{\widetilde{T}}(f)-\mu(f)\big]^2\Big] \\
&= \frac{1}{\mu(f)}\cdot\Big[p\,\mathrm{Cov}\big(g_+,f\big)^2+(1-p)\,\mathrm{Cov}\big(g_-,f\big)^2\Big] \;\le\; \max_{s\in\{+,-\}}\frac{\mathrm{Cov}\big(g_s,f\big)^2}{\mu(f)}\,, \qquad (18)
\end{aligned}$$

where $\mathrm{Cov}$ denotes covariance w.r.t $\mu$. Now observe that, since $f$ does not depend on $B_{r,\ell}$, when computing the covariance term in (18) the function $g_s$ can be replaced by $g_s^{(\ell)}$, which depends only on the spins in $\partial B_{r,\ell}$. Thus, if we can show that (17) implies

$$\mathrm{Cov}\big(g_s^{(\ell)},f\big)^2 \le c\delta\mu(f)\,\mathrm{Ent}(f) \qquad (19)$$

for some constant $c$, then by plugging (19) into (18) we will get that $\mathrm{Ent}\big[\mu_{\widetilde{T}}(f)\big] \le c\delta\,\mathrm{Ent}(f)$, as required.

To establish (19) we make use of the following technical lemma, whose proof can be found in Section 7.

**Lemma 5.4** *Let $\{\Omega,\mathcal{F},\nu\}$ be a probability space and let $f_1$ be a mean-zero random variable such that $\|f_1\|_\infty \le 1$ and $\nu[|f_1| > \delta] \le e^{-2/\delta}$ for some $\delta \in (0,1)$. Let $f_2$ be a probability density w.r.t. $\nu$, i.e. $f_2 \ge 0$ and $\nu(f_2) = 1$. Then there exists a numerical constant $c' > 0$ independent of $\nu$, $f_1$, $f_2$ and $\delta$, such that $\nu(f_1f_2)^2 \le c'\,\delta\,\mathrm{Ent}_\nu(f_2)$.*

We apply this lemma with $\nu = \mu$ and

$$f_1 = \frac{(g_s^{(\ell)}-1)}{\|g_s^{(\ell)}\|_\infty}; \quad f_2 = \frac{f}{\mu(f)},$$

to deduce $\mathrm{Cov}\big(g_s^{(\ell)},f\big)^2 \le c'\delta\|g_s^{(\ell)}\|_\infty^2\mu(f)\,\mathrm{Ent}(f)$. Noting also that $\|g_s^{(\ell)}\|_\infty \le \|g_s\|_\infty \le 1/p_{\min}$, where $p_{\min}$ was defined just before Theorem 3.4, this establishes (19) with $c = c'/p_{\min}^2$ and thus completes the proof of the theorem. $\square$

## 5.2 Proof of Theorem 5.1

In light of Theorem 5.3, to prove Theorem 5.1 it is sufficient to verify the strong concentration property (17) of the functions $g_s^{(\ell)}$ with $\delta = (\gamma\alpha)^{\ell/5}$.

In order to do this we appeal to a strong concentration of the Hamming distance under the coupling $\nu$ of $\mu^{+}_{\widetilde{T}}$ and $\mu^{-}_{\widetilde{T}}$, as defined in the proof of Claim 4.4. Recall the notation used in that claim, and notice that the Hamming distance is dominated by the size of the population in the $\ell$th generation of a specific branching process. The following tail bound can be obtained using standard techniques from the analysis of branching processes, and we defer the proof to the end of this section.

**Lemma 5.5** *Let $\alpha = \max\{\kappa b, 1\}$. Then for every $C > 0$,*

$$\Pr_\nu\Big[|\sigma-\sigma'|_{r,\ell} > C\alpha^\ell\Big] \;\le\; e^{\frac{1}{\ell+1}\left(1-\frac{C}{2e}\right)}.$$

**Corollary 5.6** *For every $C > 0$ and $s \in \{+,-\}$,*

$$\Pr_\nu\Big[\Big|g_s^{(\ell)}(\sigma)-g_s^{(\ell)}(\sigma')\Big| > C(\gamma\alpha)^\ell\Big] \;\le\; e^{\frac{1}{\ell+1}\left(1-\frac{p_{\min}C}{2e}\right)}.$$

**Proof:** It is enough to show that

$$|g_s^{(\ell)}(\sigma) - g_s^{(\ell)}(\sigma')| \le \frac{\gamma^\ell}{p_{\min}} \cdot |\sigma - \sigma'|_{r,\ell} \tag{20}$$

since we can then apply Lemma 5.5 with $C$ replaced by $p_{\min}C$. On the other hand, (20) follows from part (ii) of Claim 4.4 once we recall that $g_s^{(\ell)}(\sigma) = \mu^\sigma_{B_{r,\ell}}(g_s)$ and that $g_s$ depends only on the spin at the root, implying that $|g_s^{(\ell)}(\sigma) - g_s^{(\ell)}(\sigma')| \le \|\mu^\sigma_{B_{r,\ell}} - \mu^{\sigma'}_{B_{r,\ell}}\|_r \cdot \|g_s\|_\infty \le \gamma^\ell \, |\sigma - \sigma'|_{r,\ell} \, / p_{\min}$. □

Before we go on with the proof of Theorem 5.1, let us compare the way we used the constants $\kappa$ and $\gamma$ in the proof of Corollary 5.6 to the way we used them in the proof of Theorem 4.3. In both cases we used $\kappa$ and $\gamma$ to get bounds for coupling "down" and "up" the tree respectively. Specifically, we used $\kappa$ to deduce that the Hamming distance between the coupled configurations at the $\ell$th level is about $(\kappa b)^\ell$, and we then used $\gamma$ to bound the effect of each discrepancy at the $\ell$th level on the spin at the root (or equivalently, on $g_s^{(\ell)}$) by roughly $\gamma^\ell$. While in Theorem 4.3 it was enough that the *average* Hamming distance when coupling down the tree was bounded by $(\kappa b)^\ell$, here we need that this distance is not much larger than $(\kappa b)^\ell$ *with high probability*.

We now return to the proof of Theorem 5.1. W.l.o.g. we may assume that $\gamma\alpha \le 1$ since $\mathrm{EM}(\ell, 1)$ always holds, and also that $\gamma\alpha > 0$ since if $\gamma = 0$ then $\mathrm{EM}(\ell, 0)$ holds because then the spin at the root $r$ is independent of the rest of the configuration. Let $a = (\gamma\alpha)^{-1} \ge 1$. Recall that we wish to establish (17) with $\delta = a^{-\ell/5}$ for all large enough $\ell$. We will show only that

$$\mu\big[g_s^{(\ell)} - 1 > \delta\big] \le \frac{1}{2} e^{-2/\delta} \tag{21}$$

since the same bound on the negative tail can be achieved by an analogous argument.

We start by applying Corollary 5.6 with $C = a^{\ell/4}$ to get that, for every $\varepsilon > 0$,

$$\mu^s_{\widetilde{T}}\big[g_s^{(\ell)} - 1 > \varepsilon\big] \;\le\; \mu\big[g_s^{(\ell)} - 1 > \varepsilon - a^{-3\ell/4}\big] + A, \tag{22}$$

where $A = e^{\frac{1}{\ell+1}\left(1 - \frac{p_{\min} a^{\ell/4}}{2e}\right)}$ and we have used the fact that $\mu$ is a convex combination of $\mu^+_{\widetilde{T}}$ and $\mu^-_{\widetilde{T}}$.

Next, we notice that by definition of $g_s^{(\ell)}$,

$$\mu^s_{\widetilde{T}}\big[g_s^{(\ell)} - 1 > \varepsilon\big] \;\ge\; (1+\varepsilon)\mu\big[g_s^{(\ell)} - 1 > \varepsilon\big]. \tag{23}$$

Combining (22) and (23) we get that, for every $\varepsilon > 0$,

$$\mu\big[g_s^{(\ell)} - 1 > \varepsilon\big] \;\le\; \left(\frac{1}{1+\varepsilon}\right)\left(\mu\big[g_s^{(\ell)} - 1 > \varepsilon - a^{-3\ell/4}\big] + A\right). \tag{24}$$

This immediately yields that, for every non-negative integer $k$ and $\varepsilon > 0$,

$$\mu\big[g_s^{(\ell)} - 1 > \varepsilon + ka^{-3\ell/4}\big] \;\le\; (1+\varepsilon)^{-(k+1)} + A\left(\frac{1+\varepsilon}{\varepsilon}\right), \tag{25}$$

where we applied (24) $k+1$ times, each time increasing $\varepsilon$ by $a^{-3\ell/4}$.

Inequality (21) then follows (assuming $\ell$ is large enough) by applying (25) with $\varepsilon = a^{-\ell/4}$ and $k = \lceil a^{\ell/2} \rceil$. This concludes the proof of Theorem 5.1. □

Finally, we supply the missing proof of Lemma 5.5.

**Proof of Lemma 5.5:** First notice that, by an exponential Markov inequality, it is enough to show that $\mathrm{E}_\nu\left[e^{t|\sigma-\sigma'|_{r,\ell}}\right] \le e^{2et\alpha^\ell}$ for all $t \le (2e(\ell+1)\alpha^\ell)^{-1} \le 1$. We thus fix $t$ as above and let $D_{x,i} = \mathrm{E}_\nu\left[e^{t|\sigma-\sigma'|_{x,i}}\right]$, where $\nu$ is the coupling of $\mu^+_{\widetilde{T}_x}$ and $\mu^-_{\widetilde{T}_x}$. Note that $D_{x,i}$ can be calculated recursively as follows. The main observation is that, given a disagreement at $x$, the random variable $|\sigma-\sigma'|_{x,i}$ is the sum of the $b$ *independent* random variables $|\sigma-\sigma'|_{z,i-1}$ where $z$ ranges over the children of $x$. In turn, the random variable $e^{t|\sigma-\sigma'|_{z,i-1}}$ takes the value $D_{z,i-1}$ with probability at most $\kappa$ (the probability of a disagreement at $z$ given a disagreement at $x$) and the value 1 with the remaining probability (since $|\sigma-\sigma'|_{z,i-1} = 0$ if there is no disagreement at $z$). Thus, if we let $\delta_i = \max_x D_{x,i} - 1$, then $\delta_{i+1} \le [1+\kappa\delta_i]^b - 1 \le e^{\kappa b\delta_i} - 1 \le e^{\alpha\delta_i} - 1$. We wish to show that, for $t$ in the above range, $\delta_\ell \le 2et\alpha^\ell$, which implies $\mathrm{E}_\nu\left[e^{t|\sigma-\sigma'|_{r,\ell}}\right] \le \delta_\ell + 1 \le e^{2et\alpha^\ell}$, as required. In fact, we show by induction that $\delta_i \le 2t[\frac{\ell+1}{\ell}\cdot\alpha]^i$ for every $0 \le i \le \ell$. For the base case $i = 0$, notice that $|\sigma-\sigma'|_{x,0} = 1$ when starting from a fixed disagreement at $x$, so $\delta_0 = e^t - 1 \le 2t$ for $t$ in the given range. For $i+1 > 0$, we use the fact that $\delta_{i+1} \le e^{\alpha\delta_i} - 1 \le \frac{\alpha\delta_i}{1-\alpha\delta_i} \le \frac{\ell+1}{\ell}\cdot\alpha\delta_i$, since by the induction hypothesis $\delta_i \le \frac{1}{\alpha(\ell+1)}$ for all $0 \le i \le \ell-1$ and $t$ in the given range. $\square$

## 5.3 A crude bound on log-Sobolev via the spectral gap

In this section we state and prove a general bound on $c_{\rm sob}$ using a bound on $c_{\rm gap}$. Although we do not require this bound for the results in this paper, we believe that it may find applications in the future. We state the bound for the Ising model, but it can be easily verified that it generalizes to any nearest-neighbor spin system on a tree.

**Theorem 5.7** *For the Ising model on the $b$–ary tree, $c_{\rm sob}(\mu) = c_{\rm gap}(\mu) \times \Omega(1/\log n)$. In particular, if $c_{\rm gap}(\mu) = \Omega(1)$ then $c_{\rm sob}(\mu) = \Omega(1/\log n)$.*

It is useful to compare this bound with the well-known bound $c_{\rm sob}(\mu) = c_{\rm gap}(\mu) \times \Omega(1/n)$ (see, e.g.,[36]), which though much weaker is also more general (for example, it applies to spin systems on any graph).

Theorem 5.7 is a consequence of the following lemma.

**Lemma 5.8** *For any $\beta$ and $h$, there exists a constant $c = c(b,\beta,h)$ such that, for any $x \in T$ and all $\ell$,*

$$c_{\rm sob}(\mu^\tau_{T_x})^{-1} \le \max_{y\prec x,\eta\in\Omega^\tau_T}\{c_{\rm sob}(\mu^\eta_{T_y})^{-1}\} + c\cdot c_{\rm gap}(\mu^\tau_{T_x})^{-1}\,. \tag{26}$$

This lemma immediately implies Theorem 5.7, once we notice that $c_{\rm gap}(\mu^\eta_{T_x}) \ge c'\cdot c_{\rm gap}(\mu^\tau_T)$ for a constant $c' = c'(b,\beta,h)$ and every $x \in T$ and $\eta \in \Omega^\tau_T$, as can easily be checked.

**Proof of Lemma 5.8:** For simplicity and w.l.o.g. we will prove the recursive inequality (26) only for $T_x = T$ (the whole tree), with root $r$. Let $f$ be a non–negative function. We then write (using the entropy version of (2))

$$\mathrm{Ent}(f) = \mu\big[\mathrm{Ent}_{\widetilde{T}}(f)\big] + \mathrm{Ent}\big[\mu_{\widetilde{T}}(f)\big]\,. \tag{27}$$

Using the definition of $c_{\rm sob}$ we have

$$\begin{aligned}\mu\big[\mathrm{Ent}_{\widetilde{T}}(f)\big] &\le \max_{y\prec r,\eta\in\Omega^\tau_T}\{c_{\rm sob}(\mu^\eta_{T_y})^{-1}\}\sum_{x\in\widetilde{T}}\mu\big[\mathrm{Var}_{\{x\}}(\sqrt{f})\big]\\ &\le \max_{y\prec r,\eta\in\Omega^\tau_T}\{c_{\rm sob}(\mu^\eta_{T_y})^{-1}\}\,\mathcal{D}(\sqrt{f})\,. \end{aligned}\tag{28}$$

The second term on the r.h.s. of (27), being the entropy of a Bernoulli random variable, is bounded above by

$$\begin{aligned}\mathrm{Ent}\big[\mu_{\widetilde{T}}(f)\big] &\le \alpha \mathrm{Var}\Big(\sqrt{\mu_{\widetilde{T}}(f)}\Big) \qquad (29)\\ &\le \alpha \mathrm{Var}(\sqrt{f})\\ &\le \alpha\, c_{\mathrm{gap}}(\mu)^{-1}\,\mathcal{D}(\sqrt{f}), \qquad (30)\end{aligned}$$

where $\alpha \equiv \alpha(p)$ is a constant that depends on $p = \mu(\sigma_r = +)$; specifically $\alpha(p) = \frac{\log(p/1-p)}{2p-1}$ for $p \neq 1/2$, and $\alpha(1/2) = 1/2$ (see [36]).

Putting together (28) and (30), the expression in (27) is bounded above by

$$\Big[\max_{y\prec r,\eta\in\Omega_T^\tau}\{c_{\mathrm{sob}}(\mu_{T_y}^{\eta})^{-1}\} + \alpha\, c_{\mathrm{gap}}(\mu)^{-1}\Big]\mathcal{D}(\sqrt{f}),$$

so that from the definition of $c_{\mathrm{sob}}$ we have

$$c_{\mathrm{sob}}(\mu)^{-1} \le \max_{y\prec r,\eta\in\Omega_T^\tau}\{c_{\mathrm{sob}}(\mu_{T_y}^{\eta})^{-1}\} + \alpha\, c_{\mathrm{gap}}(\mu)^{-1}. \qquad \square$$

# 6 Extensions to other models

As we have already indicated, our techniques extend beyond the Ising model to general nearest-neighbor interaction models on trees, including those with hard constraints. In this final section we mention some of these extensions. For a fuller treatment of this material, the reader is referred to the companion paper [31].

A (*nearest neighbor*) *spin system* on a finite graph $G = (V, E)$ is specified by a finite set $S$ of spin values, a symmetric pair potential $U : S \times S \to \mathbb{R} \cup \{\infty\}$, and a singleton potential $W : S \to \mathbb{R}$. A configuration $\sigma \in S^V$ of the system assigns to each vertex (site) $v \in V$ a spin value $\sigma_v \in S$. The Gibbs distribution is given by

$$\mu(\sigma) \propto \exp\Big[-\big(\sum\nolimits_{xy\in E} U(\sigma_x,\sigma_y) + \sum\nolimits_{x\in V} W(\sigma_x)\big)\Big].$$

Thus the Ising model corresponds to the case $S = \{\pm 1\}$, and $U(s_1, s_2) = -\beta s_1 s_2$, $W(s) = -\beta h s$, where $\beta$ is the inverse temperature and $h$ is the external field. Note that setting $U(s_1, s_2) = \infty$ corresponds to a *hard constraint*, i.e., spin values $s_1, s_2$ are forbidden to be adjacent. We denote by $\Omega$ the set of all *valid* spin configurations, i.e., those for which $\mu(\sigma) > 0$.

As for the Ising model, we allow boundary conditions which fix the spin values of certain sites. We carry over our notation from the Ising model: thus, e.g., $\mu_A^\tau$ denotes the Gibbs distribution on a subset $A \subseteq V$ with boundary condition $\tau$ on $\partial A$.

The (heat-bath) Glauber dynamics extends in the obvious way to general spin systems. We first note that, as the reader may easily check, neither the spatial mixing conditions in Section 3 nor their proofs made any reference to the details of the Ising model. All of this material therefore carries over without modification to general spin systems on trees.

**Theorem 6.1** *The statements of theorems 3.2 and 3.4 hold for general nearest-neighbor spin systems on trees.*

Likewise, the machinery developed in Sections 4 and 5 for verifying the conditions VM and EM also extends to general models, though the details of the calculations are model-specific. In particular, Theorems 4.3 and 5.1 relating VM and EM to the coupling quantities $\kappa$ and $\gamma$ of Definition 4.2 still hold (with very minor modifications). Thus all we need to do is to carry out the detailed calculations of $\kappa$ and $\gamma$ for the model under consideration. We now state without proof the results of these calculations for several models of interest. For the proofs, together with further discussion and extensions, the reader is referred to the companion paper [31].

## 6.1 The hard-core model (independent sets)

In this model $S = \{0, 1\}$, and we refer to a site as *occupied* if it has spin value 1, and *unoccupied* otherwise. The potentials are

$$U(1,1) = \infty; \quad U(1,0) = U(0,0) = 1; \quad W(1) = L; \quad W(0) = 0,$$

where $L \in \mathbb{R}$. The hard constraint here means that no two adjacent sites may be occupied, so $\Omega$ can be identified with the set of all *independent sets* in $G$. Also, the aggregated potential of a valid configuration is proportional to the number of occupied sites. Hence the Gibbs distribution takes the simple form

$$\mu(\sigma) \propto \lambda^{N(\sigma)},$$

where $N(\sigma)$ is the number of occupied sites and the parameter $\lambda = \exp(-L) > 0$, which controls the density of occupation, is referred to as the "activity."

The hard-core model on a $b$-ary tree undergoes a phase transition at a critical activity $\lambda = \lambda_0 = \frac{b^b}{(b-1)^{b+1}}$ (see, e.g., [39, 23]). For $\lambda \leq \lambda_0$ there is a unique Gibbs measure regardless of the boundary condition on the leaves, while for $\lambda > \lambda_0$ there are (at least) two distinct phases, corresponding to the "odd" and "even" boundary conditions respectively. The even boundary condition is obtained by making the leaves of the tree all occupied if the depth is even, and all unoccupied otherwise. The odd boundary condition is the complement of this. (These boundary conditions are derived from the two maximum-density configurations on the infinite tree $\mathbb{T}^b$ in which alternate levels — either odd or even — are completely occupied.) For $\lambda > \lambda_0$, the probability of occupation of the root in the infinite-volume Gibbs measure differs for odd and even boundary conditions. Relatively little is known about the Glauber dynamics for the hard-core model on trees, beyond the general result of Luby and Vigoda [27, 43] which ensures a mixing time of $O(\log n)$ (after translation to our continuous time setting) when $\lambda < \frac{2}{b-1}$. This result actually holds for any graph $G$ of maximum degree $b + 1$.

Our results for the Glauber dynamics in the hard-core model mirror those given earlier for the Ising model. First, for sufficiently small activity $\lambda$ we show that both $c_{\text{gap}}$ and $c_{\text{sob}}$ are uniformly bounded away from zero for *arbitrary* boundary conditions. Second, for *even* (or, symmetrically, odd) boundary conditions, we get the same result for *all* activities $\lambda$.

**Theorem 6.2** *For the hard-core model on the $n$-vertex $b$-ary tree with boundary condition $\tau$, $c_{\text{gap}}(\mu)$ and $c_{\text{sob}}(\mu)$ are $\Omega(1)$ in both of the following situations :*

(i) *$\tau$ is arbitrary, and $\lambda \leq \max\left\{\frac{1}{\sqrt{b}-1}, \lambda_0\right\}$;*

(ii) *$\tau$ is even (or odd), and $\lambda \geq 0$ is arbitrary.*

Part (ii) of this theorem is analogous to our earlier result for the Ising model with $(+)$-boundary and zero external field at all temperatures. This is in line with the intuition that the even boundary eliminates the only bottleneck in the dynamics. Part (i) identifies a region in which the mixing time is insensitive to the boundary condition. We would expect this to hold throughout the low-activity region $\lambda \leq \lambda_0$, and indeed, by analogy with the Ising model, also in some intermediate region beyond this. Our bound in part (i) confirms this behavior: note that the quantity $\frac{1}{\sqrt{b}-1}$ exceeds $\lambda_0$ for all $b \geq 5$, and indeed for large $b$ it grows as $\frac{1}{\sqrt{b}}$ compared to the $\frac{1}{b}$ growth of $\lambda_0$. Thus for $b \geq 5$ we establish rapid mixing in a region above the critical value $\lambda_0$. To the best of our knowledge this is the first such result. (Note that the result of [27, 43] mentioned earlier establishes rapid mixing for $\lambda < \frac{2}{b-1}$, which is less than $\lambda_0$ for all $b$ and so does not even cover the whole uniqueness region.) We should also mention that our coupling analysis of $c_{\text{gap}}$ in this region has consequences for the infinite volume Gibbs measure itself, implying that when $\lambda \leq \frac{1}{\sqrt{b}-1}$ any $\mu = \lim_{T\to\infty} \mu_T^\tau$ that is the limit of finite Gibbs distributions for some boundary configuration $\tau$ is *extremal*, again a new result. We elaborate on these points in the companion paper [31].

### 6.2 The antiferromagnetic Potts model (colorings)

In this model $S = \{1, 2, \ldots, q\}$, and the potentials are $U(s_1, s_2) = \beta\delta_{s_1,s_2}$, $W(s) = 0$. This is the analog of the Ising model except that the interactions are *antiferromagnetic*, i.e., neighbors with unequal spins are favored. The most interesting case of this model is when $\beta = \infty$ (i.e., zero temperature), which introduces hard constraints. Thus if we think of the $q$ spin values as colors, $\Omega$ is the set of *proper colorings* of $G$, i.e., assignments of colors to vertices so that no two adjacent vertices receive the same color. The Gibbs distribution is uniform over proper colorings. In this model it is $q$ that provides the parameterization. For background on the model, see [8].

For colorings on the $b$-ary tree it is well known that, when $q \le b + 1$, there are multiple Gibbs measures; this follows immediately from the existence of "frozen configurations," i.e., colorings in which the color of every internal vertex is forced by the colors of the leaves (see, e.g., [8]). Recently Jonasson [21] proved that, as soon as $q \ge b+2$, the Gibbs measure is unique. Moreover, it is known that there is again an "intermediate" region that includes the value $q = b + 1$, in which the Gibbs measure, while not unique, is insensitive to "typical" boundary conditions (chosen from the free measure); see [8].

The sharpest result known for the Glauber dynamics on colorings is due to Vigoda [42], who shows that for arbitrary boundary conditions the mixing time is $O(\log n)$ provided $q > \frac{11}{6}(b + 1)$. Actually this result holds for *any* $n$-vertex graph $G$ of maximum degree $b + 1$.** Our techniques extend this rapid mixing result all the way down to the critical value $q \ge b+2$ for which uniqueness holds, with arbitrary boundary conditions. Again, our result is a consequence of the fact that the associated log-Sobolev constant is bounded below by a constant independent of $n$:

**Theorem 6.3** *For the colorings model on the $n$-vertex $b$-ary tree with $q \ge b+2$ and arbitrary boundary conditions, both $c_{\rm gap}(\mu)$ and $c_{\rm sob}(\mu)$ are $\Omega(1)$.*

### 6.3 The ferromagnetic Potts model

Here we have $S = \{1, 2, \ldots, q\}$ and potentials $U(s_1, s_2) = -\beta\delta_{s_1,s_2}$, $W(s) = 0$. This is a straightforward generalization of the (ferromagnetic) Ising model studied earlier in the paper, in which the spin at each site can take one of $q$ possible values, and the aggregated potential of any configuration depends on the number of adjacent pairs of equal spins. There are no hard constraints.

Qualitatively the behavior of this model is similar to that of the Ising model, though less is known in precise quantitative terms. Again there is a phase transition at a critical $\beta = \beta_0$, which depends on $b$ and $q$, so that for $\beta > \beta_0$ (and indeed for $\beta \ge \beta_0$ when $q > 2$) there are multiple phases. This value $\beta_0$ does not in general have a closed form, but it is known [16] that $\beta_0 < \frac{1}{2}\ln(\frac{b+q-1}{b-1})$ for all $q > 2$. (For $q = 2$, this value is exactly $\beta_0$ for the Ising model as quoted earlier.)

Using our techniques, we are able to prove the following:

**Theorem 6.4** *For the Potts model on the $n$-vertex $b$-ary tree, $c_{\rm gap}(\mu)$ and $c_{\rm sob}(\mu)$ are $\Omega(1)$ in all of the following situations:*

(i) *the boundary condition is arbitrary and $\beta < \max\left\{\beta_0, \frac{1}{2}\ln(\frac{\sqrt{b}+1}{\sqrt{b}-1})\right\}$;*

(ii) *the boundary condition is constant (e.g., all sites on the boundary have spin 1) and $\beta$ is arbitrary;*

(iii) *the boundary is free (i.e., the boundary spins are unconstrained) and $\beta < \beta_1$, where $\beta_1$ is the solution to the equation $\frac{e^{2\beta_1}-1}{e^{2\beta_1}+q-1} \cdot \frac{e^{2\beta_1}-1}{e^{2\beta_1}+1} = \frac{1}{b}$.*

**A recent sequence of papers [12, 32, 17] have reduced the required number of colors further for general graphs, under the assumption that the maximum degree is $\Omega(\log n)$; the current state of the art requires $q \ge (1+\epsilon)(b+1)$ for arbitrarily small $\epsilon > 0$ [18]. However, these results do not apply in our setting where the degree $b + 1$ is fixed.

Part (i) of this theorem shows that $c_{\rm gap}$ and $c_{\rm sob}$ are $\Omega(1)$ for arbitrary boundaries throughout the uniqueness region; also, since $\frac{1}{2}\ln(\frac{\sqrt{b}+1}{\sqrt{b}-1}) \geq \frac{1}{2}\ln(\frac{b+q-1}{b-1}) > \beta_0$ when $q \leq 2(\sqrt{b}+1)$, this result extends into the multiple phase region for many combinations of $b$ and $q$. Part (ii) of the theorem is an analog of our earlier results for the Ising model with $(+)$-boundaries at all temperatures. Part (iii) is of interest for two reasons. First, since $\beta_1 > \beta_0$ always, it exhibits a natural boundary condition under which $c_{\rm gap}$ and $c_{\rm sob}$ are $\Omega(1)$ beyond the uniqueness region (but not for arbitrary $\beta$) *for all* combinations of $b$ and $q$. Second, because of an intimate connection between the free boundary case and so-called "reconstruction problems" on trees [33] (in which the edges are noisy channels and the goal is to reconstruct a value transmitted from the root), we obtain an alternative proof of the best known value of the noise parameter under which reconstruction is impossible [34]. Indeed, a slight strengthening of part (iii) allows us to marginally improve on this threshold. Again, we spell out the details in [31].

# 7 Proofs omitted from the main text

In this final section, we supply the proofs of some technical lemmas that were omitted from the main text.

## 7.1 Proof of Lemma 3.5

The lemma in fact holds in a more general setting, where in place of $\widetilde{T_x}$ and $B_{x,\ell}$ we think of two arbitrary subsets $A, B$ such that $A \cup B = T_x$. Also, in this proof we write $\nu = \mu_{T_x}^{\eta}$ and $\mathrm{Var}$ and $\mathrm{Ent}$ for variance and entropy with respect to $\nu$. For part (i) we will show that if for any function $g$ that does not depend on $B$ we have $\mathrm{Var}[\nu_A(g)] \leq \varepsilon \cdot \mathrm{Var}(g)$, then for any function $f$,

$$\mathrm{Var}[\nu_A(f)] \leq \frac{2(1-\varepsilon)}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_B(f)] + \frac{2\varepsilon}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_A(f)].$$

Notice that by the convexity of variance we have $\mathrm{Var}(g_1+g_2) \leq 2[\mathrm{Var}(g_1)+\mathrm{Var}(g_2)]$ for any two functions $g_1, g_2$. We therefore write

$$\begin{aligned}
\mathrm{Var}[\nu_A(f)] &= \mathrm{Var}[\nu_A(f) - \nu_A(\nu_B(f)) + \nu_A(\nu_B(f))] \\
&\leq 2\mathrm{Var}[\nu_A(f - \nu_B(f))] + 2\mathrm{Var}[\nu_A(\nu_B(f))] \\
&\leq 2\mathrm{Var}[f - \nu_B(f)] + 2\varepsilon\mathrm{Var}[\nu_B(f)] \\
&= 2\nu[\mathrm{Var}_B(f)] + 2\varepsilon(\mathrm{Var}[\nu_A(f)] + \nu[\mathrm{Var}_A(f)] - \nu[\mathrm{Var}_B(f)]),
\end{aligned}$$

where we used the facts that $\mathrm{Var}[f - \nu_B(f)] = \nu[\mathrm{Var}_B(f)]$ and that $\mathrm{Var}[\nu_A(f)] + \nu[\mathrm{Var}_A(f)] = \mathrm{Var}[\nu_B(f)] + \nu[\mathrm{Var}_B(f)] = \mathrm{Var}(f)$ as in (2). We therefore conclude that $\mathrm{Var}[\nu_A(f)] \leq \frac{2(1-\varepsilon)}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_B(f)] + \frac{2\varepsilon}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_A(f)]$, as required.

We proceed to part (ii). Here we have to show that if for any non-negative function $g$ that does not depend $B$ we have $\mathrm{Ent}[\nu_A(g)] \leq \varepsilon \cdot \mathrm{Ent}(g)$, then for any non-negative function $f$,

$$\mathrm{Ent}[\nu_A(f)] \leq \frac{1}{1-\varepsilon'} \cdot \nu[\mathrm{Ent}_B(f)] + \frac{\varepsilon'}{1-\varepsilon'} \cdot \nu[\mathrm{Ent}_A(f)], \tag{31}$$

where $\varepsilon' = \sqrt{\varepsilon}/p$ and $p$ stands for the minimum non-zero probability of any configuration in $T_x \setminus A$. We will in fact show that

$$\mathrm{Ent}(f) \leq \frac{1}{1-\varepsilon'}(\nu[\mathrm{Ent}_A(f)] + \nu[\mathrm{Ent}_B(f)]), \tag{32}$$

which implies (31) since $\mathrm{Ent}[\nu_A(f)] = \mathrm{Ent}(f) - \nu[\mathrm{Ent}_A(f)]$.

Before we go on with the proof, let us review some properties of entropy. First, by definition, $\mathrm{Ent}(f) = \nu(f \log \frac{f}{\nu(f)})$ and $\nu[\mathrm{Ent}_A(f)] = \nu(f \log \frac{f}{\nu_A(f)})$. Also, by the variational characterization of entropy we have $\nu_A(f \log \frac{g}{\nu_A(g)}) \le \mathrm{Ent}_A(f)$ for all non-negative functions $f$ and $g$.

We can now proceed with the proof of (32) by writing

$$\begin{aligned}
\mathrm{Ent}(f) &= \nu\left[f \log \frac{f}{\nu_B(f)}\right] + \nu\left[f \log \frac{\nu_B(f)}{\nu_A(\nu_B(f))}\right] + \nu\left[\frac{f \log \nu_A(\nu_B(f))}{\nu(f)}\right] \\
&\le \nu\left[f \log \frac{f}{\nu_B(f)}\right] + \nu\left[f \log \frac{f}{\nu_A(f)}\right] + \nu\left[f \log \frac{\nu_A(\nu_B(f))}{\nu(f)}\right] \\
&= \nu[\mathrm{Ent}_B(f)] + \nu[\mathrm{Ent}_A(f)] + \nu\left[\nu_A(f) \log \frac{\nu_A(\nu_B(f))}{\nu(f)}\right] .
\end{aligned}$$

Therefore, (32) will follow once we show that $\nu\left[\nu_A(f) \log \frac{\nu_A(\nu_B(f))}{\nu(f)}\right] \le \varepsilon' \,\mathrm{Ent}(f)$. We use the following claim in order to get this bound.

**Claim 7.1** *Let $\mu$ be a probability measure over a space $\Omega$ where the probability of any $\sigma \in \Omega$ is either zero or at least $p$. Then for any two non-negative functions $f$ and $g$ over $\Omega$ we have*

$$\mu\left[f \log \frac{g}{\mu g}\right] \le \frac{1}{p}\sqrt{\frac{\mu(f)}{\mu(g)} \cdot \mathrm{Ent}(f) \cdot \mathrm{Ent}(g)},$$

*where* $\mathrm{Ent}$ *is taken w.r.t. to $\mu$.*

Assuming Claim 7.1, we conclude that

$$\nu\left[\nu_A(f) \log \frac{\nu_A(\nu_B(f))}{\nu(f)}\right] \le \frac{1}{p}\sqrt{\mathrm{Ent}[\nu_A(f)] \cdot \mathrm{Ent}[\nu_A(\nu_B(f))]} \le$$

$$\frac{1}{p}\sqrt{\varepsilon \cdot \mathrm{Ent}[\nu_A(f)] \cdot \mathrm{Ent}[\nu_B(f)]} \le \frac{1}{p}\sqrt{\varepsilon}\,\mathrm{Ent}(f),$$

completing the proof of Lemma 3.5. We note that, since neither $\nu_A(f)$ nor $\nu_A(\nu_B(f))$ depends on $A$, the effective probability space in the above derivation is the marginal over $T_x \setminus A$, so indeed $p$ can be taken as the minimum marginal probability of configurations restricted to $T_x \setminus A$.

It remains to prove claim 7.1. Consider two arbitrary non-negative functions $f$ and $g$. Let $\chi$ be the indicator function of the event that $g \ge \mu(g)$. Clearly, $\chi \log \frac{g}{\mu(g)} \ge 0$ while $(1-\chi)\log \frac{g}{\mu(g)} \le 0$. Also, since $\mu\left[\log \frac{g}{\mu(g)}\right] \le \log \mu\left[\frac{g}{\mu(g)}\right] = 0$ then $\mu\left[(1-\chi)\log \frac{g}{\mu(g)}\right] \le -\mu\left[\chi \log \frac{g}{\mu(g)}\right]$. Letting $f_{\max}$ and $f_{\min}$ be the maximum and minimum values of $f$ respectively over configurations with non-zero probability, we get:

$$\begin{aligned}
\mu\left[f \log \frac{g}{\mu(g)}\right] &= \mu\left[\chi f \log \frac{g}{\mu(g)}\right] + \mu\left[(1-\chi) f \log \frac{g}{\nu(g)}\right] \\
&\le f_{\max} \cdot \mu\left[\chi \log \frac{g}{\mu(g)}\right] + f_{\min} \cdot \mu\left[(1-\chi)\log \frac{g}{\mu(g)}\right] \\
&\le (f_{\max} - f_{\min}) \cdot \mu\left[\chi \log \frac{g}{\mu(g)}\right] \\
&\le \frac{1}{p} \cdot \|f - \mu(f)\|_1 \cdot \mu\left[\chi\left(\frac{g}{\mu(g)} - 1\right)\right] \\
&= \frac{1}{2p \cdot \mu(g)} \cdot \|f - \mu(f)\|_1 \cdot \|g - \mu(g)\|_1 \\
&\le \frac{1}{p}\sqrt{\frac{\mu(f)}{\mu(g)} \cdot \mathrm{Ent}(f) \cdot \mathrm{Ent}(g)},
\end{aligned}$$

where we wrote $\|\cdot\|_1$ for the $\ell_1$ norm with respect to $\mu$ and used the fact that $\|f-\mu(f)\|_1^2 \le 2\mu(f)\,\mathrm{Ent}(f)$ for any non-negative function $f$ (see, e.g., [36]). The proof of Claim 7.1 is now complete. $\square$

### 7.2 Proof of reverse direction of Theorem 3.4

In the main text we proved the forward direction of Theorem 3.4. Here we prove the reverse direction, i.e., that $\min_{x,\eta} c_{\mathrm{sob}}(\mu^{\eta}_{\tilde{T}_x}) = \Omega(1)$ implies $\mathrm{EM}(\ell, ce^{-\vartheta\ell})$ for all $\ell$, where $c = c(b,\beta,h)$ and $\vartheta = \vartheta(b,\beta,h)$ are constants independent of $\ell$. To do this, we follow the same line of reasoning as in the proof of Theorem 5.2: namely, we establish the strong concentration property of the functions $g_s^{(\ell)}$ as in Section 5.1 and then appeal to Theorem 5.3. The proof of concentration is accomplished via hypercontractivity bounds, assuming the above condition on $c_{\mathrm{sob}}$.

For a function $f$, let $\Lambda_f \subseteq T$ denote the subset of sites on whose spins $f$ depends. We then have:

**Lemma 7.2** *Let $\nu$ be any Gibbs measure on $T$, $f$ any function, and $B$ any subset that includes all sites within distance $\ell$ from $\Lambda_f$. Then there exists a constant $\vartheta'$, depending only on the degree $b$, such that*

$$\|\nu_B(f) - \nu(f)\|_q \;\le\; 3e^{-c_{\mathrm{sob}}(\nu)\vartheta'\ell}|\Lambda_f|\|f-\nu(f)\|_\infty\,,$$

*where $q = 1 + e^{c_{\mathrm{sob}}(\nu)\vartheta'\ell}$ and norms are taken w.r.t. $\nu$.*

We first assume Lemma 7.2 and complete the proof of the reverse direction of Theorem 3.4.

For simplicity, we verify EM only for the case $T_x = T$ (the whole tree), with root $r$. Recall the functions $g_s^{(\ell)}$ from Section 5.1, the fact that $g_s^{(\ell)} = \mu_{B_{r,\ell}}(g_s)$ by definition, and that $g_s$ depends only on the spin at $r$. Applying Lemma 7.2 with $\nu = \mu$, $f = g_s$, and $B = B_{r,\ell}$, together with the fact that $c_{\mathrm{sob}}(\mu) = \Omega(1)$ by hypothesis, we conclude that there exists a constant $\vartheta''$ such that

$$\|g_s^{(\ell)} - 1\|_q \le 3e^{-\vartheta''\ell}\|g_s - 1\|_\infty \le 3e^{-\vartheta''\ell}/p_{\min},$$

where $q = 1 + e^{\vartheta''\ell}$ and norms are taken w.r.t. $\mu$. Therefore, using a Markov inequality, there exist constants $\ell_0$ and $\vartheta$ such that, for all $\ell \ge \ell_0$,

$$\mu^{\pm}_{T}\big[|g_s^{(\ell)} - 1| > e^{-\vartheta\ell}\big] \;\le\; e^{-2e^{\vartheta\ell}}.$$

This establishes the strong concentration property of $g_s^{(\ell)}$ as in (17), from which EM follows by Theorem 5.3. $\square$

**Remark:** A similar claim to Lemma 7.2 was proved in [41] in the context of $\mathbb{Z}^d$; we reprove it below for completeness. The proof, as well as the fact that a $\Omega(1)$ logarithmic Sobolev constant implies $\mathrm{EM}(\ell, ce^{-\vartheta\ell})$, applies to general, finite range models on any graph of bounded degree.

**Proof of Lemma 7.2:** The proof has two main ingredients: the first is a bound on the speed at which information propagates under the Glauber dynamics, while the second is a standard relationship between $c_{\mathrm{sob}}$ and hypercontractivity bounds.

Let $P_t = e^{t\mathcal{L}}$ stand for the transition kernel at time $t$ (as discussed in Section 2) of the dynamics under consideration, reversible w.r.t. the Gibbs measure $\nu$, and let $P_t^B$ stand for the transition kernel of a modified dynamics where the spins of the sites outside the subset $B$ are fixed to their values at time zero (the sites inside $B$ being updated according to the same rule as in the original dynamics). It is well known (see, e.g., [41]) that there exists a constant $k_0$ depending only on $b$ (or on the degree of the graph in the general case) and the maximum flip rate $\max_x \|c_x\|_\infty$ (which is bounded

by 1 in the case of the heat bath dynamics) such that, for any function $f$, any $t$ and any subset $B$ that includes all sites within distance $k_0 t$ of $\Lambda_f$,

$$\|P_t f - P_t^B f\|_\infty \le 2e^{-t}|\Lambda_f|\|f\|_\infty. \tag{33}$$

Equation (33) is a manifestation of the fact that it takes at least $\frac{\ell}{k_0}$ time before the spin at a site can become sensitive to the configuration at distance $\ell$ from it.

The second ingredient we need is a hypercontractivity bound. From Gross's integration lemma (see, e.g., [1]), we have $\|P_t f\|_q \le \|f\|_2$ for any mean-zero function $f$, any $t$, and $2 \le q \le 1 + e^{c_{\rm sob} t}$, where $c_{\rm sob} = c_{\rm sob}(\nu)$. Adding to this the fact that $c_{\rm gap} > c_{\rm sob}$, we may write

$$\|P_t f\|_q = \|P_{t/2}(P_{t/2} f)\|_q \le \|P_{t/2} f\|_2 \le e^{-c_{\rm gap} t/2}\|f\|_2 \le e^{-c_{\rm sob} t/2}\|f\|_2, \tag{34}$$

where $q = 1 + e^{c_{\rm sob} t/2}$ and we used the fact that $c_{\rm gap}$ bounds the rate of decay of the $L^2$ norm.

We now conclude the proof of Lemma 7.2 as follows. Without loss of generality, consider an arbitrary function $f$ with $\nu(f) = 0$. Let $\ell$ be arbitrary, and $B$ be a subset that includes all sites within distance $\ell$ of $\Lambda_f$. Then, for $t = \ell/k_0$ and $q = 1 + e^{c_{\rm sob} t/2}$, we have

$$\begin{aligned}
\|\nu_B(f)\|_q &= \|\nu_B(P_t^B f)\|_q \\
&\le \|P_t^B f\|_q \\
&\le \|P_t^B f - P_t f\|_q + \|P_t f\|_q \\
&\le 2e^{-t}|\Lambda_f|\|f\|_\infty + e^{-c_{\rm sob} t/2}\|f\|_2 \\
&\le 3e^{-c_{\rm sob}\vartheta\ell}|\Lambda_f|\|f\|_\infty\,,
\end{aligned}$$

taking the constant $\vartheta = 1/2k_0$ (and using the fact that $c_{\rm sob} \le 1$). □

### 7.3 Proof of Lemma 5.4

We split our analysis of $\nu(f_1 f_2)^2$ into three cases:

(a) $\mathrm{Ent}_\nu(f_2) \ge \frac{1}{\delta}$;

(b) $\delta < \mathrm{Ent}_\nu(f_2) < \frac{1}{\delta}$;

(c) $\mathrm{Ent}_\nu(f_2) \le \delta$.

*Case (a)*. We simply bound

$$\nu(f_1 f_2)^2 \le \|f_1\|_\infty^2 \nu(f_2)^2 \le 1 \le \delta\ \mathrm{Ent}_\nu(f_2)\,.$$

*Case (b)*. We use the *entropy inequality* (see, e.g., [1]), which states that for any $t > 0$,

$$\nu(f_1 f_2) \le \frac{1}{t}\log\nu(e^{tf_1}) + \frac{1}{t}\mathrm{Ent}_\nu(f_2)\,. \tag{35}$$

We choose the free parameter $t$ in (35) equal to $\sqrt{\mathrm{Ent}_\nu(f_2)/\delta}$. Notice that, by construction, $1 < t < \delta^{-1}$. Using the assumption $\nu(|f_1| > \delta) \le e^{-2/\delta}$ together with $\|f_1\|_\infty \le 1$, we get

$$\begin{aligned}
\nu(f_1 f_2)^2 &\le \Big[\frac{1}{t}\log\big(e^{t\delta} + e^{t-2/\delta}\big) + \sqrt{\delta\,\mathrm{Ent}_\nu(f_2)}\,\Big]^2 \\
&\le \Big[c_1\,\delta + \sqrt{\delta\,\mathrm{Ent}_\nu(f_2)}\,\Big]^2 \le c_2\,\delta\,\mathrm{Ent}_\nu(f_2)
\end{aligned}$$

for suitable numerical constants $c_1, c_2$.

*Case (c)*. Again we use the entropy inequality with $t = \sqrt{\mathrm{Ent}_\nu(f_2)/\delta} \le 1$, but we now simply bound the Laplace transform $\nu(e^{tf_1})$ by a Taylor expansion (in $t$) up to second order:

$$\begin{aligned}\frac{1}{t}\log\nu(e^{tf_1}) &\le \frac{1}{t}\log\Big(1 + e\frac{t^2}{2}\nu(f_1^2)\Big) \le e\frac{t}{2}\big[\,\delta^2 + e^{-2/\delta}\,\big]\\ &= \frac{1}{2}e\big[\,\delta^2 + e^{-2/\delta}\,\big]\sqrt{\mathrm{Ent}_\nu(f_2)/\delta},\end{aligned}$$

which by (35) implies

$$\nu(f_1f_2)^2 \le \Big[\frac{e}{2\sqrt{\delta}}\big(\,\delta^2 + e^{-2/\delta}\,\big) + \sqrt{\delta}\,\Big]^2 \mathrm{Ent}_\nu(f_2) \le c_3\,\delta\,\mathrm{Ent}_\nu(f_2)$$

for another numerical constant $c_3$. □

## References


[1] C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO and G. SCHEFFER, "Sur les inégalités de Sobolev logarithmiques," *Société Mathématique de France*, 2000.

[2] R.J. BAXTER, *Exactly solved models in statistical mechanics*, Academic Press, London, 1982.

[3] N. BERGER, C. KENYON, E. MOSSEL and Y. PERES, "Glauber dynamics on trees and hyperbolic graphs," preprint (2003). Preliminary version: C. KENYON, E. MOSSEL and Y. PERES, "Glauber dynamics on trees and hyperbolic graphs," *Proc. 42nd IEEE Symposium on Foundations of Computer Science* (2001), pp. 568–578.

[4] L. BERTINI, N. CANCRINI and F. CESI, "The spectral gap for a Glauber-type dynamics in a continuous gas," *Ann. Inst. H. Poincaré Probab. Statist.* **38** (2002), pp. 91–108.

[5] P. BLEHER, J. RUIZ, R.H. SCHONMANN, S. SHLOSMAN and V. ZAGREBNOV, "Rigidity of the critical phases on a Cayley tree," *Moscow Mathematical Journal* **1** (2001), pp. 345–363.

[6] P. BLEHER, J. RUIZ and V. ZAGREBNOV, "On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice," *Journal of Statistical Physics* **79** (1995), pp. 473–482.

[7] T. BODINEAU and F. MARTINELLI, "Some new results on the kinetic Ising model in a pure phase," *Journal of Statistical Physics* **109** (1), 2002.

[8] G. BRIGHTWELL and P. WINKLER, "Random colorings of a Cayley tree," *Contemporary combinatorics*, Bolyai Society Mathematical Studies **10**, János Bolyai Math. Soc., Budapest, 2002, pp. 247–276.

[9] F. CESI, "Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields," *Probability Theory and Related Fields* **120** (2001), pp. 569–584.

[10] J. T. CHAYES, L. CHAYES, J.P. SETHNA and D.J. THOULESS, "A mean field spin glass with short-range interactions" *Communications in Mathematical Physics* **106** (1986), pp. 41–89.

[11] R. DOBRUSHIN, R. KOTECKÝ and S. SHLOSMAN, "Wulff Construction. A Global Shape From Local Interaction" *Translation of Math. Monographs*, AMS **104** (1992).

[12] M. DYER and A. FRIEZE, "Randomly colouring graphs with lower bounds on girth and maximum degree," *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, 2001, pp. 579–587.

[13] W. EVANS, C. KENYON, Y. PERES and L.J. SCHULMAN, "Broadcasting on trees and the Ising model," *Annals of Applied Probability* **10** (2000), pp. 410–433.

[14] D. FISHER and D. HUSE, "Dynamics of droplet fluctuations in pure and random Ising systems," *Physics Review B* **35** (13), 1987.

[15] H.-O. GEORGII, *Gibbs measures and phase transitions*, de Gruyter Studies in Mathematics **9**, Walter de Gruyter & Co., Berlin, 1988.

[16] O. HÄGGSTRÖM, "The random-cluster model on a homogeneous tree," *Probability Theory and Related Fields* **104** (1996), pp. 231–253.

[17] T.P. HAYES, "Randomly coloring graphs with girth five," *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, 2003, pp. 269–278.

[18] T.P. HAYES and E. VIGODA, "A non-Markovian coupling for randomly sampling colorings," to appear in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[19] D. IOFFE, "A note on the extremality of the disordered state for the Ising model on the Bethe lattice," *Letters in Mathematical Physics* **37** (1996), pp. 137–143.

[20] D. IOFFE, "Extremality of the disordered state for the Ising model on general trees," *Progress in Probability* **40** (1996), pp. 3–14.

[21] J. JONASSON "Uniqueness of uniform random colorings of regular trees," *Statistics & Probability Letters* **57** (2002), pp. 243–248.

[22] J. JONASSON and J.E. STEIF, "Amenability and phase transition in the Ising model," *Journal of Theoretical Probability* **12** (1999), pp. 549–559.

[23] F.P. KELLY, "Stochastic models of computer communication systems," *Journal of the Royal Statistical Society* B **47** (1985), pp. 379–395.

[24] R. LYONS, *Phase transitions on non amenable graphs*, J.Math.Phys **41**, pp. 1099–1127, 2000.

[25] M. LEDOUX, "The concentration of measure phenomenon," *Mathematical Surveys and Monographs* **89**, American Mathematical Society, 1981.

[26] S.L. LU and H.T. YAU, "Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics" *Comm. Math. Phys.* **156** (1993), pp. 399–433.

[27] M. LUBY and E. VIGODA, "Fast convergence of the Glauber dynamics for sampling independent sets," *Random Structures & Algorithms* **15** (1999), pp. 229–241.

[28] F. MARTINELLI, "Lectures on Glauber dynamics for discrete spin models," *Lectures on Probability Theory and Statistics (Saint-Flour, 1997)*, Lecture notes in Mathematics **1717**, pp. 93–191, Springer, Berlin, 1998.

[29] F. MARTINELLI and E. OLIVIERI, "Approach to equilibrium of Glauber dynamics in the one phase region I: The attractive case," *Comm. Math. Phys.* **161** (1994), pp. 447–486.

[30] F. MARTINELLI, E. OLIVIERI and R. SCHONMANN, "For 2-D lattice spin systems weak mixing implies strong mixing," *Comm. Math. Phys.* **165** (1994), pp. 33–47.

[31] F. MARTINELLI, A. SINCLAIR and D. WEITZ, "Fast mixing for independent sets, colorings and other models on trees," preprint, 2003.

[32] M. MOLLOY, "The Glauber dynamics on colorings of a graph with high girth and maximum degree," *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, 2002, pp. 91–98.

[33] E. MOSSEL, "Survey: Information flow on trees," Preprint, October 2002, to appear in DIMACS volume *Graphs, Morphisms and Statistical Physics*.

[34] E. MOSSEL and Y. PERES, "Information flow on trees," *Annals of Applied Probability*, 2003, to appear.

[35] Y. Peres and P. Winkler, personal communication.

[36] L. Saloff-Coste, "Lectures on finite Markov chains," *Lectures on probability theory and statistics (Saint-Flour, 1996)*, Lecture notes in Mathematics **1665**, pp. 301–413, Springer, Berlin, 1997.

[37] R.H. Schonmann and N.I. Tanaka, "Lack of monotonicity in ferromagnetic Ising model phase diagrams," *Annals of Applied Probability* **8** (1998), pp. 234–245.

[38] B. Simon, *The statistical mechanics of lattice gases, Vol. I*, Princeton Series in Physics, Princeton University Press, Princeton, NJ, 1993.

[39] F. Spitzer, "Markov random fields on an infinite tree," *Annals of Probability* **3** (1975), pp. 387–398.

[40] D.W. Stroock and B. Zegarlinski, "The logarithmic Sobolev inequality for discrete spin systems on a lattice," *Comm. Math. Phys.* **149** (1992), pp. 175–194.

[41] D.W. Stroock and B. Zegarlinski, "On the ergodic properties of Glauber dynamics," *J. Statist. Phys.* **81** (1995), pp. 1007–1019.

[42] E. Vigoda, "Improved bounds for sampling colorings," *Journal of Mathematical Physics* **41** (2000), pp. 1555–1569.

[43] E. Vigoda, "A note on the Glauber dynamics for sampling independent sets," Electronic Journal of Combinatorics, Volume 8(1), 2001.